Ontology-Based Roles Association Networks for Visualizing Trends in Political Debate

Troels Andreasen[1], Henning Christiansen[1], and Mads Kæmsgaard Eberholst[2]

[1] Research Group PLIS: Programming, Logic and Intelligent Systems
[2] Research Group: Communication, Journalism and Social Change
Dept. of Communication, Business and Information Technologies
Roskilde University, Denmark
Email: {troels,henning,makaeb}@ruc.dk

Abstract. Online resources, large data repositories and streaming social network messages embed plenitudes of interesting knowledge, often of associative nature. A specific communicative context, such as the political debate in a given country, has groupings of actors, with changing attitudes and stances towards each other and external, real or invented, threats and opportunities. A new form of associative network is introduced, that integrate flexible ontologies for complex contexts of roles and hierarchies with a labelled association structure representing observed strengths and attitudes. Twitter messages from the political landscape in Denmark up to the general election 2015 are used as a both current and relevant illustrative case.

1 Motivation and Background

@bansoe And so it begins....#fv15[1]

This tweet, like many others on May 27th 2015, initiated the Danish national elections on the social media platform Twitter. This paper outlines a graphical representational formalism that allows modeling how the election and its main actors have represented various subjects in the debate on Twitter, and with what sentiment. Twitter in Denmark is a primarily political (as opposed to a social) platform [1, 2] and as both media and politicians is very aware of this, the platform can generate a large amount of data very fast when there is an election. This is primarily due to the nature of the Twitter platform making it highly suited for breaking news [3] and thus rapidly developing events and attitudes. This is especially true when an election is in process, where tweets from political actors mostly are used to campaigning, spreading information or self promotion [4]. For many end users and indeed both media professionals and politicians, making grasps of such a large amount of data can be cumbersome bordering impossible. Therefore a modeling approach to visualize and understand some of the mechanics and contents of Twitter data in a political and electoral context is very much needed. For this paper we used the open source tool Yourtwapperkeeper

[1] #fv for Danish "Folketingsvalg", the election for the Danish national parliament.

to harvest all tweets from the start of the election period in Denmark to the end of the campaign. This tool has proven a viable way of gathering Twitter data which due to API limitations can be hard to do [5]. We selected the appropriate electoral hashtags that appeared to be trending within the first 24 hours. These were #dkpol (Danish general politics) and a number of hashtags all referring to the election (#fv15, #valg15, #valg2015, #drdinstemme, #tv2valg, #ft15, #dkvalg, #ftvalg15 and #dkmedier). This gave a sample of 260,000 tweets on which our methodology is tested.

Co-occurrence networks and semantic networks based on co-occurrences have been used in many areas and in different shapes. We propose an integration with an ontological concept lattice to a general model for representing complex contexts including large sets of actors with many different roles and organizational groupings – as is the case in politics. The concept lattice allows to zoom in and out, from single individual to, say, the known members of a given political party, and vice-versa, and to express asymmetric and annotated relationships, anticipating different visualizations.

2 ORAN: Ontology-Based Roles Association Networks

Ontology-Based Roles Association Networks, for short ORANs, are association networks for context representation, whose nodes are concepts taken from a *concept lattice* serving as an ontology. Concepts may be extended with *roles*, so for example, concept "Smith" may refer to any message relating somehow to a single person, whereas predicating with the role "author", forming concept "author:Smith" may refer to messages written by that person. Formally, we assume finite sets of *atomic concepts* $\mathcal{A}$, and roles $\mathcal{R}$; a *concept* is of one of the forms, a or $r{:}a$ where $a \in \mathcal{A}$, $r \in \mathcal{R}$. The concept lattice is a partially ordered set of concepts $\langle \mathcal{C}, < \rangle$ where $<$ is transitive, and if $r{:}a \in \mathcal{C}$, then $a \in \mathcal{C}$ and $r_a < a$. The relation $<$ is read "more specialized than". A concept $c \in \mathcal{C}$ has an *extension*, written $[\![c]\!]$, which is a subset of some universe $\mathcal{U}$ (e.g., of messages) such that, for any c, c' with $c < c'$, it holds that $[\![c]\!] \subseteq [\![c']\!]$. The lattice is not fully generative, as not all roles can apply to all atomic concepts, e.g., refugees are often mentioned (concept about:refugees) but do not have voices as authors.

An *ORAN* is a directed or undirected, labelled graph, described as a triplet of a concepts lattice, a subset of its atomic concepts and one or more association types, $\langle\langle \mathcal{C}, < \rangle, A, T\rangle$. An *association type* is of the form $P_1 \rightarrow P_2$ in case of a directed network and $P_1 - P_2$ for an undirected one, where P_1, P_2 are *generic concepts* of form $c \in \mathcal{C}$, $*$ or $r{:}*$, where $r \in \mathcal{R}$. The association type generates the actual set of nodes N in the graph and the edges. Concept c indicates $c \in N$, $*$ that $A \subseteq N$ and $r{:}*$ that $\{r{:}a \mid a \in A\} \cap \mathcal{C} \subseteq N$. All edges whose endpoints $\in N$ match the association type are included, so e.g., author:$* \rightarrow *$ may create, among others, author: MembersOfDanishPeoplesParty $\rightarrow$ refuges. When drawing these edges, they are typically shown between the given atomic concepts.

The edges are referred to as *associations*, and each edge $c_1 \rightarrow c_2$ or $c_1 - c_2$ has one or more *labels*, one of which is its *association degree*, also called *strength*, defined for the directed, resp. undirected case, as follows.

$$d(c_1 \to c_2) = \begin{cases} \frac{|[\![c_1]\!] \cap [\![c_2]\!]|}{|[\![c_1]\!]|}; & \text{if } |[\![c_1]\!]| > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$d(c_1 - c_2) = \begin{cases} \frac{|[\![c_1]\!] \cap [\![c_2]\!]|}{|[\![c_1]\!] \cup [\![c_2]\!]|} & \text{if } |[\![c_1]\!] \cup [\![c_2]\!]| > 0 \\ 0 & \text{otherwise} \end{cases}$$

Here $|set|$ refers to the number of elements in *set*. For our present application, the extension of a concept is given as the set of messages that refers to that concept (possible taking roles into account), and thus $d(c_1 \to c_2)$ is the fraction of the messages referring to c_1 that also refers to c_2. For example, $d(\text{author}\colon \text{Hansen} \to \text{Jensen})$ is the proportion of Hansen's messages that refer to Jensen (in some way or another).

Other possible labels attached to edges $c_1 \to c_2$ can be defined from properties of $[\![c_1]\!]$ and $[\![c_2]\!]$. In our example, we use a *sentiment* that measures a degree of positiveness/negativeness for sets of tweets.

Extracting Concept Extensions and Rules from Twitter Data

Twitter messages include sender identification and other metainformation, and the text may includes nametags (when specific Twitter users are mentioned), hashtags as well as ordinary words. These can all be mapped to concepts, specifically sender identification into author roles, and references in the text into mentions (e.g., indicated by role 'about:'). This is not an easy task, as these tags are not a controlled vocabulary, and the same person may be referred to by a nametag, different hashtags, the name written in different ways and with spelling errors. To create concepts corresponding to members (given by nametags and user id) of specific political parties, we have used an external resource, provided by the online resource http://www.twittervalgkortet.dk. Our sentiment analysis is based on a list of common Danish words, each having a "joy index" learned by statistical means,[2] ranging from -5 for the most negative to $+5$ for the most positive. We added our own stemming algorithm to extend covering and take the average over words identified in each tweet, followed by averaging over a set of tweets.

Visualizing Ontology-Based Roles Association Networks

A vast collection of software of software tools are available for presenting associative information graphically, e.g., d3.js[3] which is an impressive java script library suited for application development. For our own experiments, we stored ORANs extracted from Twitter messages in a relational database, from which

[2] This list has been prepared by Finn Årup Nielsen, Technical University of Denmark, referred is referred to in a student project [6].

[3] See http://d3js.org.

information were exported to the NodeXL tool,[4] which is a plug-in to Excel. Typical views are complete networks, showing all possible associations between the selected nodes, and so-called ego-centric views, with a single concept c in the middle, captured by the association type $c \to *$ (or $c \to r{:}*$).

Association degrees are typically shown as thickness of the edges, and other labels can be shown, e.g., by the colour of edges or be available in pop-up menus. A zooming facility is under development which can be accessed by clicking the mouse over the nodes of the network: zooming amounts to replace one or several nodes in A by other, above or below wrt. $<$.

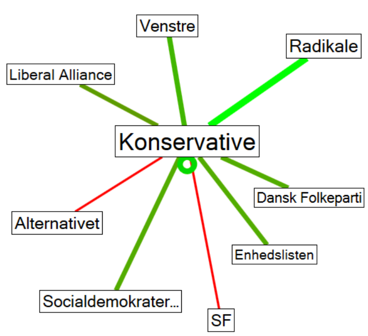
When data are sufficiently dense and subject of fast, dynamic changes as is the case for tweets about politics in an election, it is interesting also to take into account the time dimension: for the chosen view, images are produced for the tweets from each day, and these images are put together to an animation. This may give a very clear illustration of the development of attitudes.

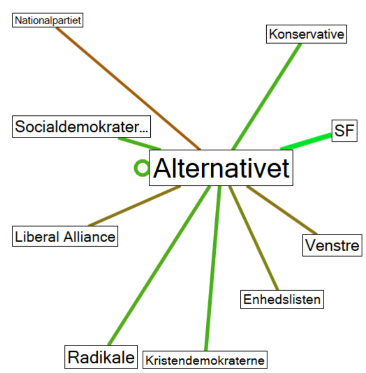
3 The Danish General Election 2015

The following two views have nodes corresponding to political parties, measured for the entire period, and illustrate the overall sentiments in how the parties *Alternativet* and *Konservative* writes about the other parties. This corresponds to the ego-centric association patterns

$$\text{author:Konservative} \to * \qquad \text{and} \qquad \text{author:Alternativet} \to *.$$

The sentiment is shown by colour, ranging from red for the most negative, and bright green for the most positive. The strengths of the associations are given by their thickness.





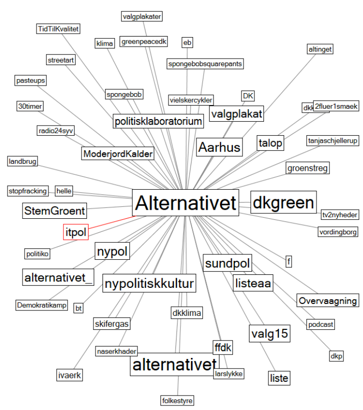
It appears that the Conservative party is very negative in its attitude towards Alternativet and SF (Socialist Peoples' Party) and basically kind for the rest. Their best friends seem to "Radikale" and then follows "Venstre".[5] We see that

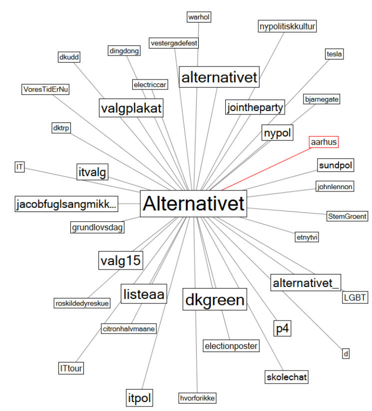
[4] See http://research.microsoft.com/en-us/projects/nodexl/.

[5] For understanding: "Radikale" is not very radical anymore, and "Venstre" literally, "left", is a clearly right wing party. The Conservative party is, however, conservative.

Alternativet is generally kind, except towards the most leftist and the most rightist parties.

Next, we illustrate a dynamic, potentially animated view, here with snapshots for two different weeks of the distribution of the most common topics that the party Alternativet is tweeting about. I.e., the nodes are $\{\text{author:Alternativet}\} \cup$ *HotTopics*, where *HotTopics* are generated from the actual hashtag set. The association pattern is as above, but with a changed set of atomic concepts.[6] Strength is now indicated by the size of the topics boxes.





4 Conclusion and Future Work

We demonstrated a novel model for context representation, based on an ontology suited for complex systems of actors with different roles and hierarchical and overlapping groupings (multiple inheritance is inherent in the model). We showed how knowledge can be extracted from social media data streams, mapped into such a model and visualized accordingly. Our test example was Twitter messages with political contents from the announcement of and the hectic days until the Danish national election were held June 18, 2015. This study will be continued into the period after the election, more data will be acquired and analyzed in order to give a complete, zoomable and animated picture of the communication around an election. As part of this, improved sentiment analysis and topic extraction will be developed, taking into account synonymy, tweet idioms and typos, as well as semantic-pragmatic considerations.

There is a lot of recent work on summarizing and visualizing data from Twitter in sophisticated ways, e.g., [7–10], but a tight coupling to an ontology that allows easily shifting view of the data, as we have described, is not common. The work of [8] traces trends in Twitter data over time using a co-occurence of topics based model. A generic approach for monitoring message streams such as Twitter, for recognizing interesting events automatically, is described by [11],

[6] The selection of the associations with strengths higher than some threshold is not part of the ORAN formalism, but a facility in the prototype implementation.

which may be interesting to combine with the present approach. Automatic of learning associative networks from Twitter is used by [12]. More detailed emotions are used in a study that correlates Twitter sentiments to socio-economic phenomena [13]. This group has also identified a high correlation of sentiment in Twitter messages with stock prices [14].

References

1. Buch, R.: Special issue on: Sociale nyhedsmedier og journalistik. Journalistica - Tidsskrift for Forskning I Journalistik (1) (2011)
2. Schrøder, K., Nielsen, R.: Danskernes brug af digitale medier og nyheder i 2014. Research report from *Center for Magt, Medier og Kommunikation, CBIT, Roskilde Universitet.* https://www.ruc.dk/fileadmin/assets/cbit/MMK/Danskernes_brug_af_digitale_medier_og_nyheder_i_2014.pdf (2014)
3. Elkjær, B.: Breaking news! journalistik på twitter. http://medieblogger.dk/wp-content/uploads/Bagger-sagen-Om-twitter-journalistik.pdf (2009)
4. Ausserhofer, J., Maireder, A.: National politics on twitter. Information, Communication & Society **16**(3) (2013) 291–314
5. Bruns, A.: How long is a tweet? mapping dynamic conversation networks on twitter using gawk and gephi. Information, Communication & Society **15**(2) (2012) 1323–1351
6. Engels, M.F., Jørgensen, S.B., Holm, H.L.: Medieoverblik.dk. http://medieoverblik.dk/wp-content/uploads/s103182_s103201_s103214_Project-AssignmentB.pdf (2014) Student report, Technical University of Denmark.
7. Arzucan Ozgur, Burak Cetin, H.B.: Co-occurrence network of Reuters news. International Journal of Modern Physics C **19**(5) (2011) 689–702
8. Malik, S., Smith, A., Hawes, T., Papadatos, P., Li, J., Dunne, C., Shneiderman, B.: Topicflow: visualizing topic alignment of twitter data over time. In Rokne, J.G., Faloutsos, C., eds.: Advances in Social Networks Analysis and Mining 2013, ASONAM '13, Niagara, ON, Canada - August 25 - 29, 2013, ACM (2013) 720–726
9. Ghanem, T.M., Magdy, A., Musleh, M., Ghani, S., Mokbel, M.F.: Viscat: spatio-temporal visualization and aggregation of categorical attributes in twitter data. In Huang, Y., Schneider, M., Gertz, M., Krumm, J., Sankaranarayanan, J., eds.: Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Dallas/Fort Worth, TX, USA, November 4-7, 2014, ACM (2014) 537–540
10. Guille, A., Favre, C.: Event detection, tracking, and visualization in twitter: a mention-anomaly-based approach. Social Netw. Analys. Mining **5**(1) (2015) 18:1–18:18
11. Andreasen, T., Christiansen, H., Have, C.T.: Tracing shifts in emotions in streaming social network data. Submitted (2015)
12. Diemert, E., Vandelle, G.: Unsupervised query categorization using automatically-built concept graphs. In Quemada, J., León, G., Maarek, Y.S., Nejdl, W., eds.: Proceedings of the 18th International Conference on World Wide Web, WWW 2009, Madrid, Spain, April 20-24, 2009, ACM (2009) 461–470
13. Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. Journal of Computational Science **2**(1) (2011) 1–8
14. Bollen, J., Pepe, A., Mao, H.: Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. (2011) 450–453