

---

## Wikimaps: dynamic maps of knowledge

---

### Reto Kleeb\*

MIT Center for Collective Intelligence,  
5 Cambridge Center,  
Cambridge, 02139, MA, USA  
E-mail: reto@techbase.ch  
\*Corresponding author

### Peter A. Gloor

MIT Center for Collective Intelligence,  
5 Cambridge Center,  
Cambridge, 02139, MA, USA  
E-mail: pgloor@mit.edu

### Keiichi Nemoto

Fuji Xerox Co., Ltd.,  
6-1 Minatomirai, Nishi-ku, Yokohama-shi,  
Kanagawa 220-8668, Japan  
E-mail: keiichi.nemoto@fujixerox.co.jp

### Michael Henninger

UAS Northwestern Switzerland,  
Institute of 4D Technologies,  
Steinackerstrasse 5, CH-5210 Windisch, Switzerland  
E-mail: Michael.henninger@fhnw.ch

**Abstract:** We introduce Wikimaps, a tool to create a dynamic map of knowledge from Wikipedia contents. Wikimaps visualise the evolution of links over time between articles in different subject areas. This visualisation allows users to learn about the context a subject is embedded in, and offers them the opportunity to explore related topics that might not have been obvious. Watching a Wikimap movie permits users to observe the evolution of a topic over time. We also introduce two static variants of Wikimaps that focus on particular aspects of Wikipedia: latest news and people pages. ‘Who-works-with-whom-on-Wikipedia’ (W5) links between two articles are constructed if the same editor has worked on both articles. W5 links are an excellent way to create maps of the most recent news. PeopleMaps only include links between Wikipedia pages about ‘living people’. PeopleMaps in different-language Wikipedias illustrate the difference in emphasis on politics, entertainment, arts and sports in different cultures.

**Keywords:** Wikipedia; scientific visualisation; intercultural comparison; knowledge mapping.

**Reference** to this paper should be made as follows: Kleeb, R., Gloor, P.A., Nemoto, K. and Henninger, M. (2012) 'Wikimaps: dynamic maps of knowledge', *Int. J. Organisational Design and Engineering*, Vol. 2, No. 2, pp.204–224.

**Biographical notes:** Reto Kleeb received his Master in Computer Science from Northeastern University in 2011 and was a visiting student at MIT in summer 2011.

Peter A. Gloor is a Research Scientist at the MIT Center for Collective Intelligence.

Keiichi Nemoto is a Researcher at Fuji Xerox. He was a visiting scholar at the MIT Center for Collective Intelligence in 2010–2011.

Michael Henninger is a Research Assistant at the Institute for 4D Technologies at University of Applied Sciences, Northwestern Switzerland.

This paper is a revised and expanded version of a paper entitled 'Wikimaps: dynamic maps of knowledge', presented at 3rd International Conference on COINs, University of Applied Sciences Northwestern Switzerland, HyperWerk Institute for Postindustrial Design, Basel, Switzerland, 8–10 September 2011.

---

## 1 Introduction

People have long predicted the demise of traditional news media and the rise of the citizen journalists. Various initiatives have tried to create new media outlets on the web, blog and twitter powered by creative swarms of hobby journalists – but none of them has been a breakthrough success so far. In this paper, we introduce a novel system to track the evolution of latest news and cultural differences based on Wikipedia.

In a series of earlier projects, we have analysed collaboration among Wikipedia authors when creating new Wikipedia articles, for example studying how they collaborate as collaborative innovation networks (COINs) in different cultures (Nemoto and Gloor, 2010). Wikipedia does not only provide the digital world with a vast amount of high quality information, the meta data about authors and the detailed information about every single contribution opens new opportunities to investigate the processes that lie behind the creation of the content as well as the relations between knowledge domains. Wikipedia has a great amount of hidden information in the metadata surrounding the content pages: what – the edits of the pages, when – edits over time, who – which authors edited the pages, and how – which links to other Wikipedia pages and outside web pages are embedded in the content.

Combining existing information with metadata allows the discovery of new, previously hidden information. Novel visualisations built on this information will greatly enhance the knowledge experience provided by Wikipedia by offering networks of ideas and concept maps and 'current hot topics' by intellectual field, societal issue, culture, etc. Even more, studying and comparing idea networks over time and in different-language Wikipedias will contribute towards better understanding of different cultures through what these cultures are passionate about. In addition, by constructing co-authorship networks where links between Wikipedia articles are drawn based on the same person

editing the different articles, we will be able to locate domain experts as well as trusted arbitrators, which will give us yet another dimension of analysing Wikipedia content and weighing the importance of edits and articles.

In the project described in this paper, we focus on creating dynamic visualisations of article-link networks. To achieve informative and visually appealing knowledge maps of the data, we are addressing two sub-tasks: fetching data to generate graphs of Wikipedia articles and their relations as close to real time as possible as well as the visualisation of these graphs over chosen time-periods. Besides the challenges related to gathering of the data, we also work on an additional research question: how do we find the most prominent, interesting and relevant pages about a chosen subject over time. While the search function provided by Wikipedia offers a good starting point to find the most relevant pages, it does not return a more fine-grained semantic network of relevant articles. In this paper we introduce a system that combines different metrics and metadata from Wikipedia.

While this project is in an early stage, it builds on three years of research in our group, studying Wikipedia co-authorship and edit networks, as well as a vast body of research in Wikipedia authorship and content by a vibrant global research community. We already have been able to show that Wikipedians form long-lasting collaboration networks resulting in high quality output. We are convinced that including these and other results will help us in building a new lens into the knowledge of mankind captured in Wikipedia, providing – we hope – yet another stepping stone towards more creativity and innovation.

## **2 Related work**

There is a lot of previous work regarding the network properties of Wikipedia as well as the visualisation of information based on Wikipedia. Chan et al. (2009) created Vispedia, a mash-up interface to visualise Wikipedia data in geographic maps, timelines and scatterplots. Weld et al. (2008) combined WordNet with the structured infoboxes of Wikipedia articles to construct ontologies. Folksoviz (Lee et al., 2008) use the information available in Wikipedia to create semantic graphs of delicious (<http://delicious.com/>) tags. (Holloway et al., 2007) use the Wikipedia article categories to construct a network map of all pages in Wikipedia, using colour-coding to optically highlight different dimensions such as edit times, or author activity. Viegas et al. (2004) have built a well-known system to create visualisations of Wikipedia editing activities named HistoryFlow and Chromogram.

While all these systems generate fascinating insights into the work of the Wikipedia authors and the relations of the content on the platform, none of the systems have combined the three elements: search, display of the results as a semantic network as well as the evolution of this network over time. We will now describe our approach combining the three elements.

## **3 WikiMaps system overview**

The WikiMaps system is divided into two distinct parts. The first part of the application fetches the information either from the Wikipedia servers or if applicable, from a local

cache. It then uses a configurable chain of filtering methods to reduce the number of possible candidates. Our goal was to create a system which would rely exclusively on the Wikipedia API as a single source of information – not weaving in any external sources such as, e.g., search results from the Google or Bing API – and that the only user input would be the definition of the search term.

The output of this fetch and filter component is a list of nodes and edges in JSON format that then can be used by the second component. This component renders the network as an interactive graph in any modern browser.

### *3.1 Initial data collection*

There are numerous studies that focus on the structure of Wikipedia as well as the interaction of the article authors (Ganjisaffar et al., 2009; Nunes et al., 2008; Yeh et al., 2009). The majority of these projects use static datasets; these are datasets that only represent a single point in time. The most current version is approximately 29.5 GB (uncompressed, English Wikipedia, 3.6 million articles, April 2011) and can easily be handled with out-of-the-box hardware resources. The problem with these datasets is that they do not include any historical information and are therefore not usable for studies that track changes in the content over time. Another problem with these dumps for a specific date is that they are only provided roughly once a month. This resolution would not allow the (close to) real-time tracking of ongoing developments. In addition to these ‘most-recent’ dumps, the Wikipedia foundation also provides complete dumps that contain the complete history. These complete dumps are unfortunately of large size, the files are currently around 5 TB which makes their handling rather impractical.

To mitigate this issue, we decided to develop a system that would not depend on a local database that contains the complete information, but implement a system that gradually builds a local cache that only contains the information that is absolutely necessary. This trade-off allowed us to build a system that minimises the amount of required local storage while still having the possibility to benefit from a local cache for repeated queries. The fetcher minimises the number of required calls to the Wikipedia HTTP API.

### *3.2 Definition of the used graphs*

Inter-Wiki Links (Wikipedia article pages to Wikipedia article pages) are represented as edges in the graph. These links are parsed directly from the Wikimedia mark-up source of the articles. If we were only interested in the most recent revision of an article, the Wikipedia API would provide a method to simply extract all internal links. Unfortunately, the same functionality is not provided of older revisions of articles.

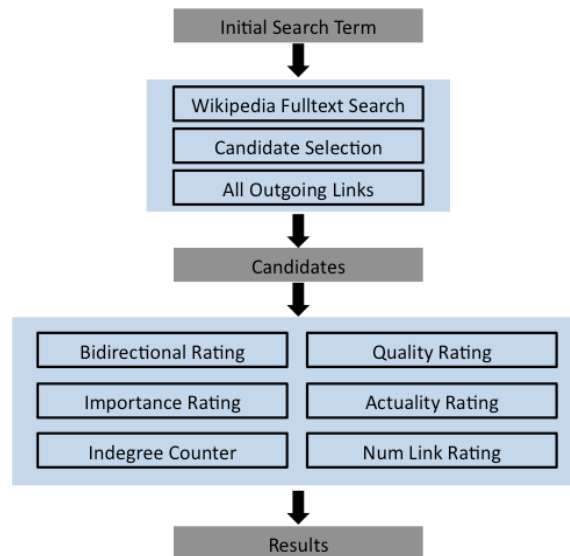
Frequently, there are multiple links from article A to a different article B. The decision whether these multiple links are set within the article lies within the responsibility of the author. The guidelines mention “as a rule of thumb only the first occurrence of a term should be linked” [[http://en.wikipedia.org/wiki/Wikipedia:Manual\\_of\\_Style\\_\(linking\)](http://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style_(linking))]. This rule is however not applied consistently and there are many cases where authors set links to the same article multiple times. Considering this situation, using the number of links between pages to assign a weight factor to the graph-edges would skew the data in a meaningless way.

### 3.3 Filtering the data

Previous research has shown that the usage of various graph metrics leads to distinctly different results that widely differ in quality (Fuehres et al., 2012). To solve this problem and lessen the dependence on individual metrics, our approach is based on a weighted combination of multiple elements. The following data sources and metrics have been used: Wikipedia fulltext search, article activity (number of edits over time period), (local) indegree, shortest path distance to search term, bidirectional links.

The algorithm that drives WikiMaps and creates the networks of article nodes is divided into two steps (Figure 1). In the first step, the system creates a list of potential candidates and in the following step, the algorithm reduces the number of nodes. The goal of this second step is to boil down the list of potential candidates to a subset that only contains the most interesting and relevant articles in regard to the original search term.

**Figure 1** Link rating and selection of Wikipedia pages to be displayed (see online version for colours)



### 3.4 Collecting all candidates

The initial step of the algorithm is the collection of all the articles that potentially could be related to the given search term. For this step, the system first executes a Wikipedia full-text search (using the HTTP API) where the top  $n$  (typically about 20) articles are stored. The system then collects all outgoing links to other Wikipedia articles for each of the articles. Depending on the popularity of the subject the set typically contains between 1,000 and 2,500 articles.

The final step for the generation of the output is to filter the nodes for each date that is desired for the animation. This algorithm reduces the list of potential articles to the most interesting subset, by combining the best candidates from three different pools. The three pools are defined as follow:

- The top nodes that are mutually connected to the initial search-term T (there is a link from the article T to another article Y as well as a link from Y back to T). The candidates in this pool are ordered by their local indegree (only links within the network of candidates C are considered).
- The top nodes that are connected (path in the directed graph) to the initial search-term T, ordered by shortest-path-distance.
- The last step ensures that the network contains enough nodes in case where the first two pools did not generate enough viable candidates. This is typically the case for subjects that are not as well connected on Wikipedia. The system adds more article-nodes until the limit for the total number of nodes in the graph is reached. If the number of articles is already larger than 50 this last ‘filling’ step is not carried out. These nodes are the top nodes of the article set, ordered by their (local) indegree. The level of graph-connectivity to the search term T is not relevant for the selection of these nodes. This step has the further advantage that it potentially adds an element of serendipity to the resulting graph. The consideration of nodes that have weaker ties increases the chances of discovering articles that might be less popular but still very interesting to find.

Based on well-known user interface norms (Shneiderman and Aris, 2006) we set the limit for the total number of nodes to be shown simultaneously in the graph to less or equal than 50.

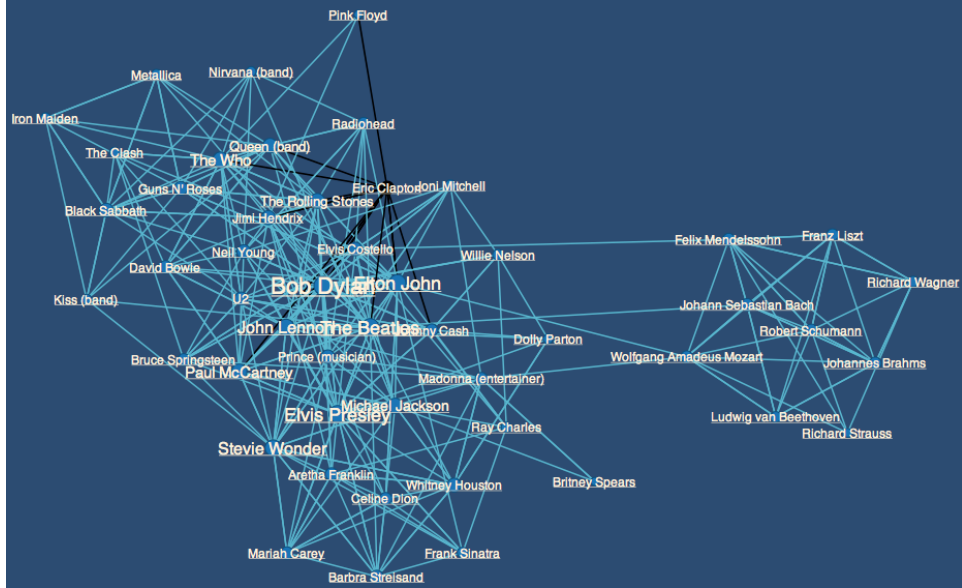
### 3.5 Visualising the graph

The final graph is visualised by the JavaScript/SVG-based Protovis (<http://protovis.org>) framework. The framework simplifies the arrangement of nodes and edges according to a force-based layout (Fruchterman and Rheingold, 1991). The toolkit further allows the computation of the node radii-based on a node’s indegree. This indegree and radii then influence the (virtual) electrical charge of the nodes, so that ‘bigger’ and more important nodes appear more central in the final rendering.

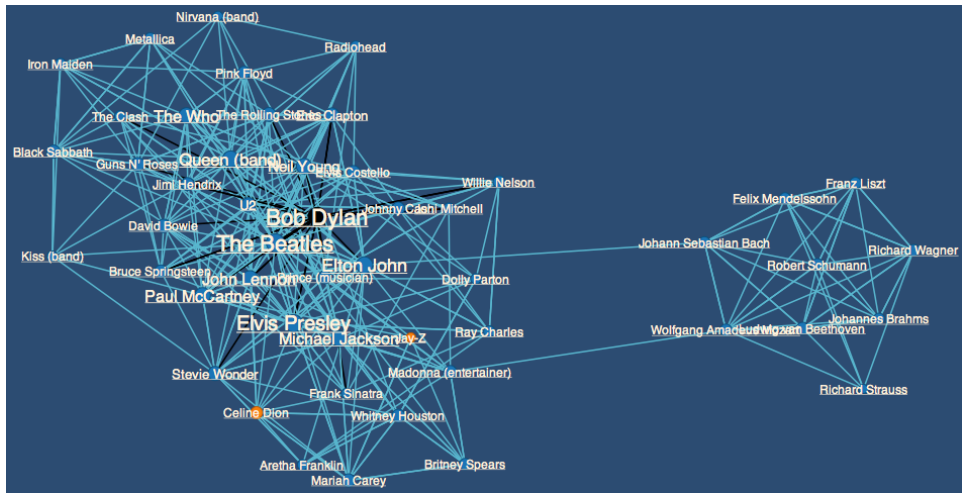
To display how the graph evolves over time, we need to repeat the algorithm for each of the desired dates. The initial set of articles remains the same but the contents of the articles are fetched individually for each date. The collection of individual graphs is then incorporated into an interactive animation that allows anybody with a web browser to track the evolution of a given subject over the course of time. The algorithm that we chose (Fruchterman and Rheingold, 1991) to arrange the nodes makes it simple to visually recognise clusters that are related to a specific sub-topic and to discover the increase or decrease of the connectivity of clusters.

The following example displays a case of a decrease in connectivity between two subject clusters on Wikipedia. In this specific case, the semantic network was built based on all musician-related categories (Figure 2 and Figure 3). The network clearly separates into a cluster of more modern pop and rock-oriented music and a second cluster containing important classical musicians. Besides the clear separation of the two clusters, we can also recognise a significant decrease of connectivity over time between the two clusters. The articles in the pop/rock cluster on the left in 2009 used to link more to the classical musicians on the right than in 2011. Similar work by Gloor et al. (2004) has been done to visualise patterns in human interaction.

**Figure 2** Important musicians on Wikipedia (February 2009) (see online version for colours)

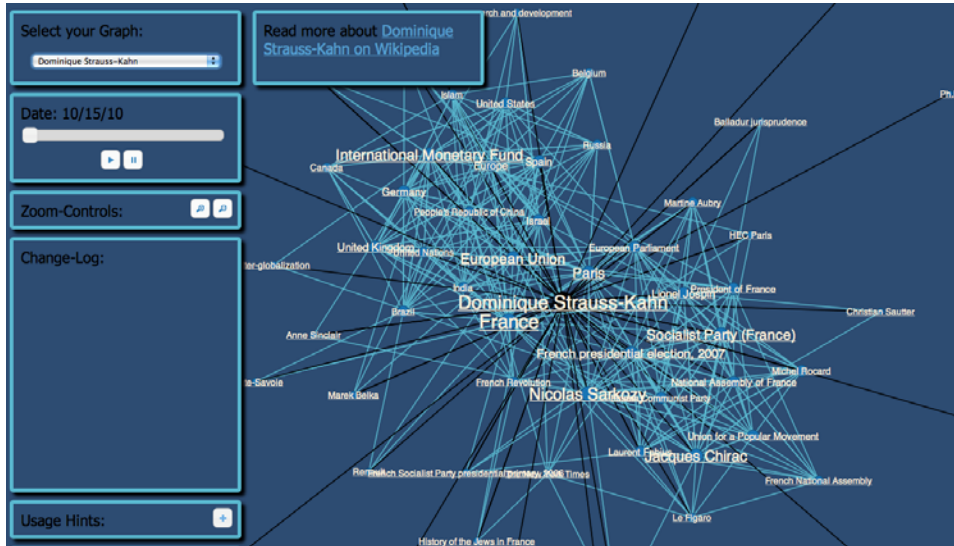


**Figure 3** Important musicians on Wikipedia (June 2011) (see online version for colours)



Our experimentation with the combination and the weighing of different matrices and graph properties lead to highly relevant semantic networks for the given subjects. The browser-based visualisation of these networks allows interested people to simply get an overview of related subjects as well as their evolution over a period of time. An alpha version of WikiMaps is available online at <http://www.ickn.org/wikimaps>. The complete functionality is available as part of Condor (Gloor and Zhao 2004).

**Figure 4** Wikimaps visualisation for the article ‘Dominique Strauss-Kahn’ (October 2010)  
(see online version for colours)



**Figure 5** Wikimaps visualisation for the article ‘Dominique Strauss-Kahn’ (July 2011)  
(see online version for colours)

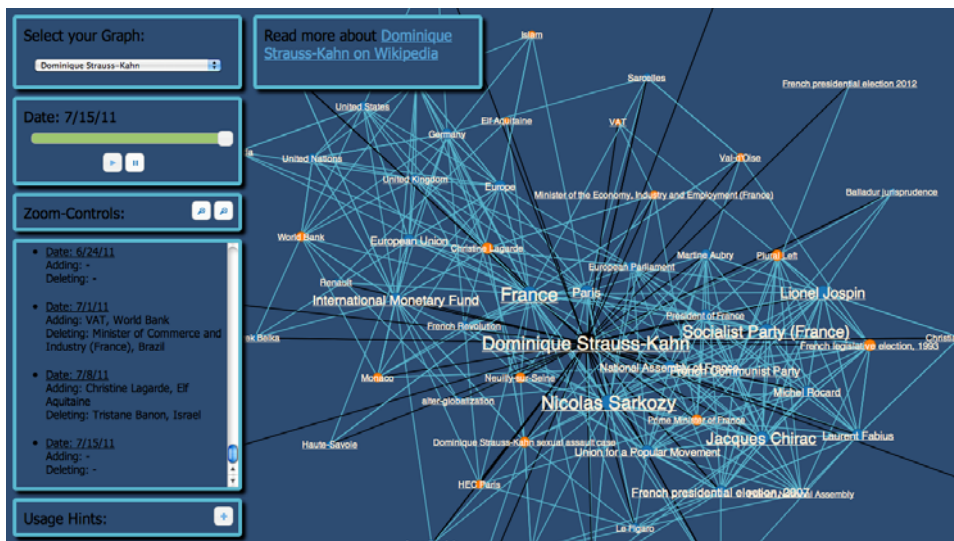


Figure 4 and Figure 5 show the complete user-interface of the alpha version of the WikiMaps application. It allows the selection of a pre-calculated animation and then offers the user the possibility to zoom in an out to explore the articles that belong to the network. This concrete example shows the progress of the graph that is based on the initial search term ‘Dominique Strauss-Kahn’. Mr. Kahn, the former president of the International Monetary Fund (IMF) was arrested in New York and charged with the sexual assault of a housekeeper. The timeframe of the animation ranges from October 2010 until July 2011 (Mr. Kahn was arrested in May 2011). The first frame of the



animation simply shows the network that Dominique Strauss-Kahn is embedded in, it is easy to detect the two clusters surrounding the main topic: IMF is in the top left corner and the articles related to French politics are in the bottom right.

The second frame shows the semantic network after the enormous media hype surrounding the arrest and the events that directly followed. The nodes that were added since the first frame are highlighted in orange. The most obvious change is the fact that the article that is dedicated to the events that took place in May 2011 (Dominique Strauss-Kahn sexual assault case) appears in the graph. We can further recognise that the cluster that is dedicated to IMF activities lost some weight, while the second major cluster related to French politics is now more densely connected.

#### **4 Reflection on Wikimaps**

Wikipedia has rather different link structures for different types of articles: articles covering current events such as presidential elections or sports have a rich linking structure; the most heavily linked pages however are about countries. Articles about basic concepts such as ‘happiness’ or ‘creativity’ have a much more sparse link network. This means that different strategies for constructing Wikimaps have to be applied to construct visually appealing maps. In the current version implemented in Condor (Fuehres et al., 2012) this is a semi-manual process where the parameters can be chosen accordingly. In future versions this could be automated by incorporating information about topic categories.

After this first overall description of WikiMaps, the next two sections will describe two specific variants of WikiMaps focused on latest events and people pages.

#### **5 Identifying news through W5-links**

To identify latest news on Wikipedia, we introduce a new type of link, the ‘who-works-with-whom-on Wikipedia’ (W5) link. We build a semantic network of concepts by constructing a link between two Wikipedia articles if the same author has worked on both articles. As the complete Wikipedia includes millions of article, drawing a whole map of Wikipedia in one step is too much. Instead we employed a ‘snowball sampling’ method, which allows us to draw a partial map by selecting a start article or editor. For our first experiment, we used the article about ‘Wikipedia’ as the starting point. We collected the top ten editors based on the number of edits on this article, and then we gathered the top ten articles of each editor. We repeated this steps recursively up to 3° of separation from the start point. Restricting this analysis to a certain period of time (e.g., one month starting January 1, 2010) permits us to obtain a temporal W5 map from this start point. Applying this process repeatedly we calculated 11 snapshots of one month each from January 2010 to November 2010. Each node corresponds to an article in Wikipedia. We draw an edge between articles A and B if there are at least two editors who made edits both on article A and article B.

Figure 6 and Figure 7 show our results. Each map was drawn by Gephi, an opensource graph visualisation tool (<http://www.gephi.org>). The size of the article title was determined by the undirected PageRank score of the W5 network.

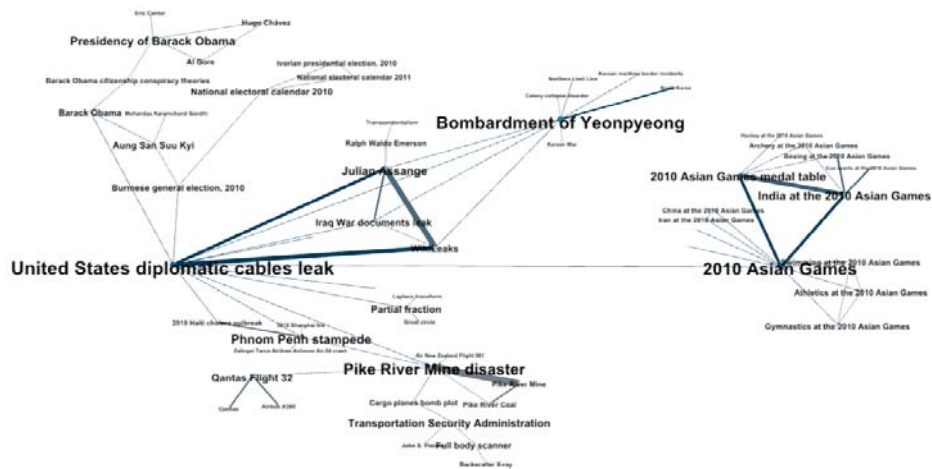


The major topics (based on PageRank Score) for each month are shown. Surprisingly they reflect the major news item of the month (Table 1).

**Table 1** Major topics of monthly W5 maps in 2010

<i>Time</i>	<i>Major topic</i>
January 2010	2010 Haiti earthquake
February 2010	2010 Winter Olympics
March 2010	2010 Polish Air Force Tu-154 crash
April 2010	Telephone (song)
May 2010	Gaza flotilla raid
June 2010	2010 FIFA World Cup
July 2010	2010 FIFA World Cup
August 2010	2010 Israel-Lebanon border clash
September 2010	2010 Atlantic hurricane season
October 2010	2010 Atlantic hurricane season
November 2010	US diplomatic cables leak

**Figure 8** Zooming on the W5 map for November displays Wikileaks and the bombardment of Yeonpyeong (see online version for colours)



The W5 map of November 2010 (Figure 8) illustrates zooming in on the Wikileaks case and the Bombardment of Yeonpyeong. It seems that groups of similarly minded Wikipedians tend to aggregate around a set of articles on similar topics they are most interested in. These groups of ‘latest news buffs’ assume the role of trend indicators, with the most important news attracting most editors. The US diplomatic cables leak is strongly connected to WikiLeaks and Julian Assange, which makes perfect sense because both of them are part of the WikiLeaks dispute. The article on the bombardment of Yeonpyeong has many ‘shared editor’ edges with the WikiLeaks cluster while there are almost no edges to the 2010 Asian Games cluster. This means that “birds of a feather flock together”: Wikipedians working on the bombardment of Yeonpyeong are interested

in other politically sensitive topics such as the Wikileaks issue, they are not interested however in sports in Asia.

Our preliminary investigation suggests that looking at Wikipedia through the W5 map might be a new way to identify latest news. We find the news of the world even if we start from a neutral article such as the one about ‘Wikipedia’. The swarm of Wikipedians seems to be a perfect group of coolhunters and citizen journalists to report latest news on politics, celebrities and sports.

## **6 WikiPeopleMaps**

Our third way of analysing Wikipedia focuses on links between Wikipedia pages about people. Articles about people in Wikipedia include many social cues about a particular person. By using this information, we create a social network of people listed in the Wikipedia article. We only look at links from one people page to another people page. The links between these pages convey a rich social context, such as friendship, animosity, communication, or having participated in the same events, for example. The meaning of a link includes relationship information about a person, for instance, with whom they have worked as well as basic demographic information such as when and where the person was born.

In the English Wikipedia there is a category called ‘living people’ which includes articles of people currently alive. This category has 478,382 articles of living people as of February 2011. To construct the social network of living people, we collected each article’s contents and extracted the internal links to other living people articles. Wikipedia provides an APIs to get a list of internal links of each article, but this does not work well in this case because the API does not distinguish between links to and from template pages and links added manually by Wikipedians which are the ones we are interested in. We therefore wrote our own link-collection crawler.

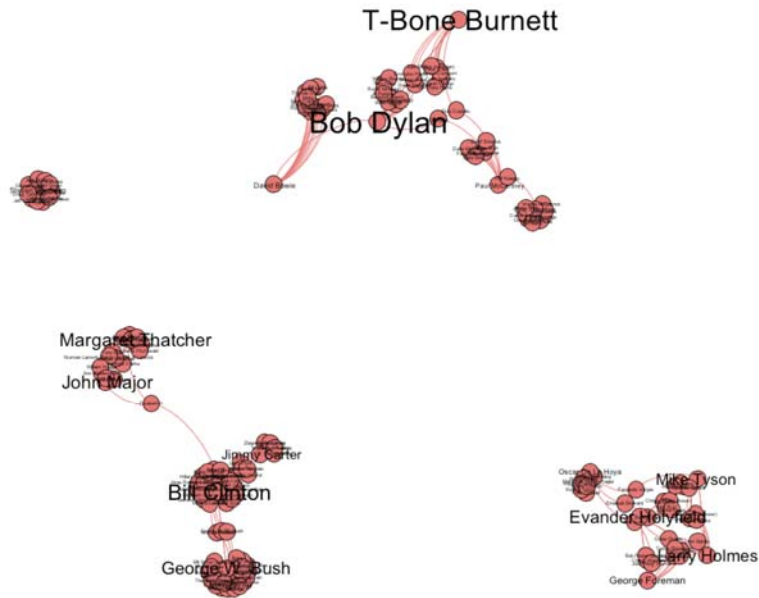
Similarly to the page rank algorithm that the Google search engine provides, this linking structure gives us a clue about which person is more prominent from the Wikipedians’ perspective. As metric of importance we took the ratio between in-degree and out-degree (Wasserman and Faust, 1994) because having a higher in-degree might be a proxy for the power or importance of the person described in the article. Because the size of the original map was too huge (over half a million articles) to be visualised, we only included articles in the visualisation having more than a predefined number of incoming links (i.e., higher in-degree) and those peripheral articles that have a direct link to the most prominent articles.

To draw a temporal map of the ‘living people network’, we collected snapshots of the contents of the living people articles at different points in time. The threshold is determined by the number of nodes and edges that can be visualised. Therefore, it varies at different points in time.

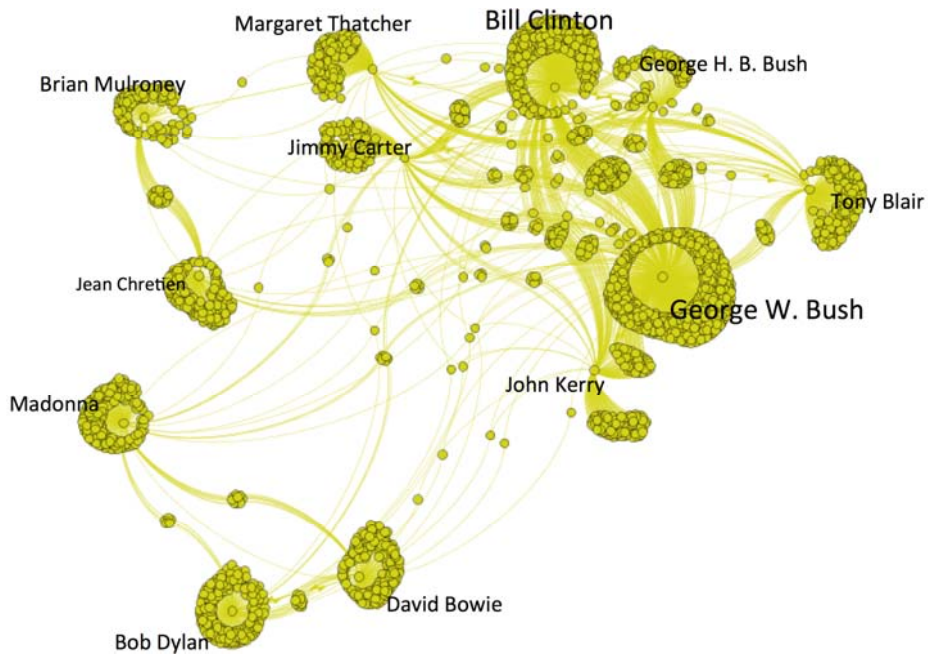
Figures 9–12 show the living people network of the English Wikipedia from 2003 to 2011. In the very early stage of Wikipedia, the living people network is unconnected and has three components, namely politicians, musicians and boxers. In 2006, politicians, especially the (former) presidents or prime ministers, are the most prominent articles about living people. The article about Barack Obama does not yet appear on the 2006 map; however it becomes one of the most prominent articles in 2009. This is not surprising because Barack Obama was elected to the US presidency in January 2009.

This result suggests that the living people network reflects the social condition and the prominence of a person in almost real-time.

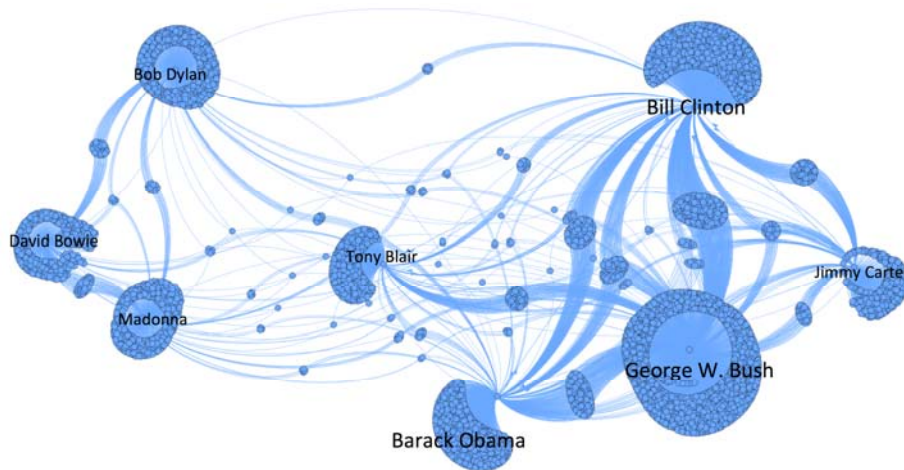
**Figure 9** Living people network in the English Wikipedia in 2003 (threshold: 10 by in-degree) (see online version for colours) (see online version for colours)



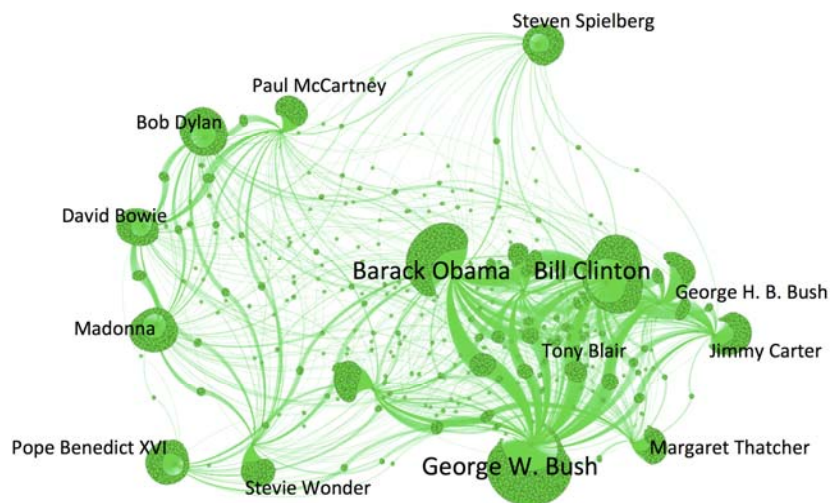
**Figure 10** Living people network in the English Wikipedia in 2006 (threshold: 150 by in-degree) (see online version for colours)



**Figure 11** Living people network in the English Wikipedia in 2009 (threshold: 500 by in-degree)  
(see online version for colours)



**Figure 12** Living people network in the English Wikipedia in 2011 (threshold: 500 by in-degree)  
(see online version for colours)



We also visualised living people networks in the different-language Wikipedias. Table 2 shows the number of articles belonging to the living people category except for the German and French Wikipedia. These two Wikipedias do not have the living people category, rather they include a series of categories, which list people by birth in every year (e.g., 1,900 births) and death in every year (e.g., 2,000 deaths). To collect a set of living people articles, we collected the articles of people who were born after 1880 and not dead until 2010.

As Table 2 illustrates, the English, German, French and Japanese Wikipedias are not only the largest Wikipedias by number of articles, they have also the highest fraction of people pages. We speculate that similar to a ‘who’s who’, the more prominent a Wikipedia becomes, the higher the incentive of ego-boasting people to put up their own

people page to promote themselves. We have studied this behaviour in earlier work (Iba et al., 2009).

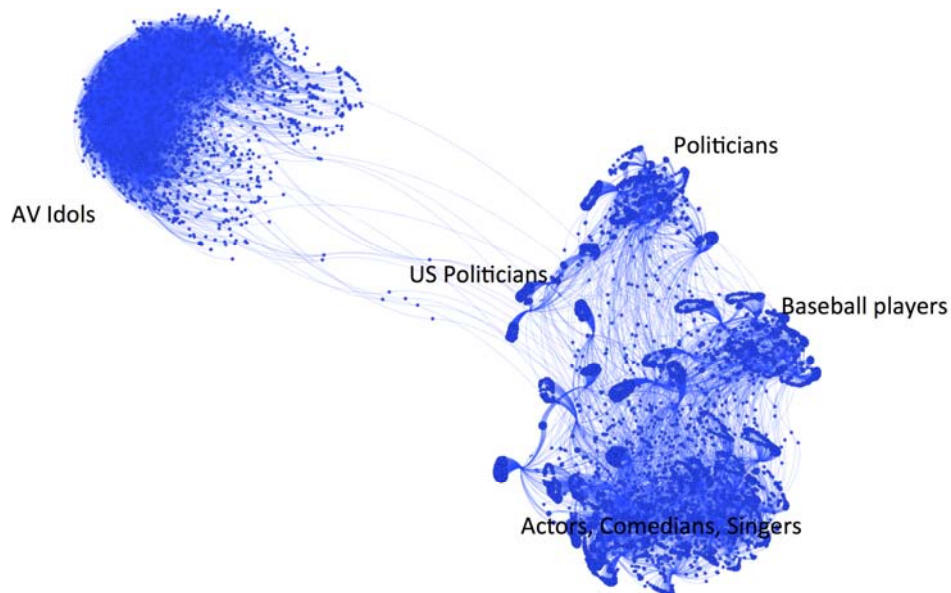
**Table 2** Number of ‘living people pages’ in the different-language Wikipedias

<i>Language</i>	<i># of living people articles (Feb. 2011)</i>	<i># of articles (Feb. 2011)</i>	<i>Ratio</i>
English	478,382	3,700,000	12.9%
German	194,043	1,200,000	16.2%
French	126,053	1,100,000	11.5%
Japanese	102,082	742,000	13.8%
Spanish	41,827	728,000	5.7%
Chinese	23,963	339,000	7.1%
Korean	5,379	158,000	3.4%

Next, we constructed the living people network for March 2011 for each of the countries listed in Table 2. We found that each different-language Wikipedia has another perspective on the prominence of people. In some Wikipedias, politicians are most prominent among living people. In other Wikipedias, entertainers, musicians and athletes are most prominent rather than politicians. The most recent US presidents have highest centrality in all Wikipedias, except in the Chinese one.

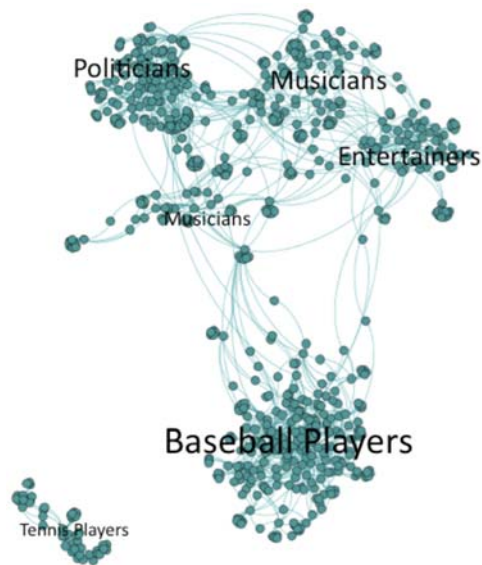
In the Japanese Wikipedia (Figure 13), there is one big independent cluster that includes a lot of AV (adult video) idols articles. The other clusters can be divided into four different parts, US politicians, Japanese politicians, baseball players and actors, comedians and singers. This suggests that people in politics are not as prominent in the Japanese Wikipedia community, but entertainers are important. This is very different from the English, German and French Wikipedia.

**Figure 13** Japanese living people network in 2011 (threshold: 150 by in-degree) (see online version for colours)

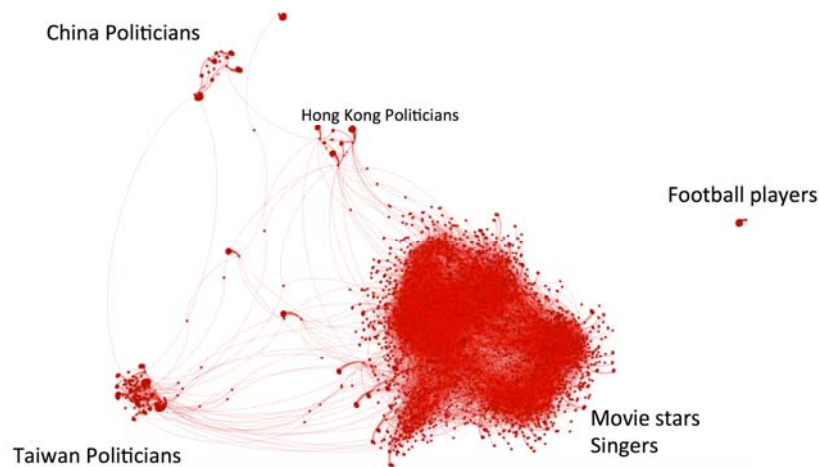


Looking at other Asian-language Wikipedias like the Chinese and Korean ones, we found quite a similar structure in the maps. In the Chinese Wikipedia (Figure 15), the one big cluster consists of movie stars and singers. Three small and disconnected clusters represent politicians in China, Hongkong and Taiwan. The wide distance between the Chinese and Taiwanese politicians reflects political reality. In the Korean Wikipedia (Figure 14), entertainers and musicians are also dominant in the network. Besides, there is a large group of baseball players – baseball is very popular in Korea, and a smaller cluster of tennis players – tennis is not very popular in Korea. The final cluster consists of politicians.

**Figure 14** Korean living people network in 2011 (threshold: 10 by in-degree) (see online version for colours)

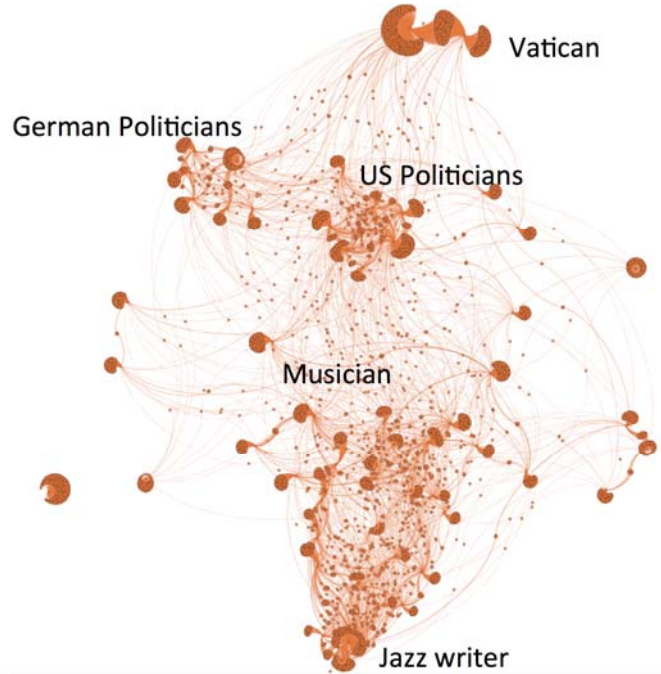


**Figure 15** Chinese living people network in 2011 (threshold: 30 by in-degree) (see online version for colours)

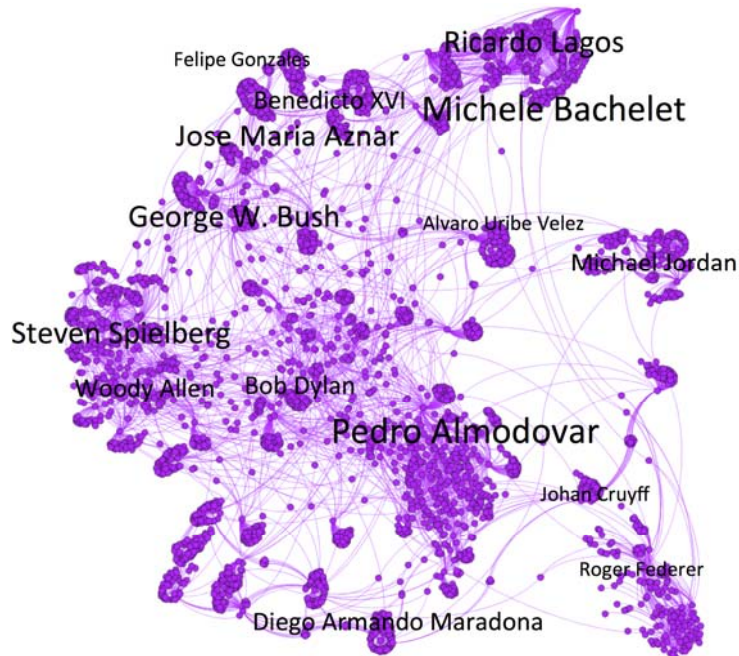




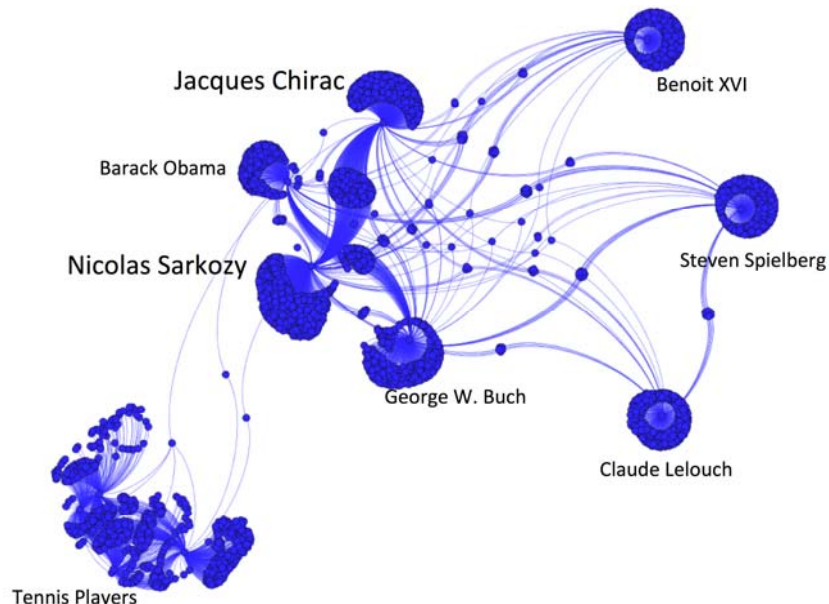
**Figure 16** German living people network in 2011 (threshold: 150 by in-degree) (see online version for colours)



**Figure 17** Spanish living people network in 2011 (threshold: 50 by in-degree) (see online version for colours)



**Figure 18** French living people network in 2011 (threshold: 250 by in-degree) (see online version for colours)



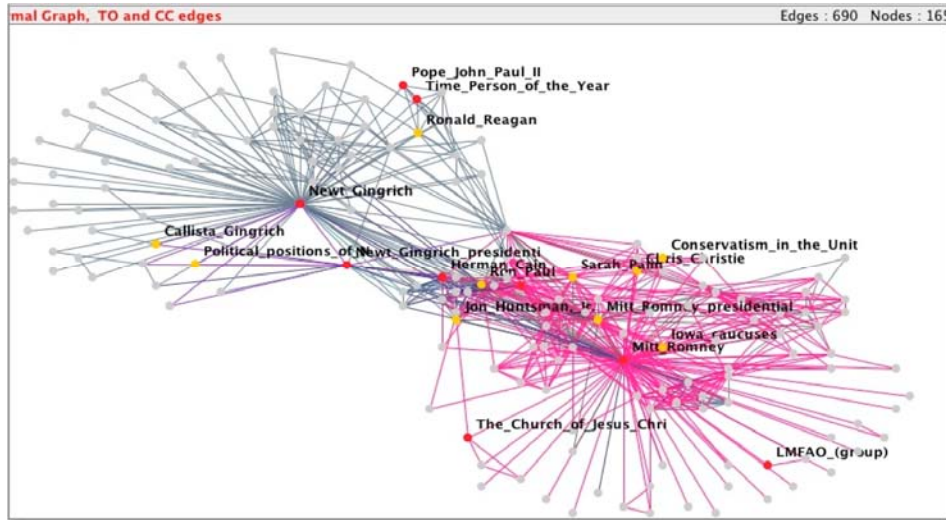
In the German Wikipedia people map (Figure 16), the top three recognisable clusters consist of German politicians, US politicians, and religious articles – explained through the fact that the current Pope is German, and the rest consists of musicians. German Wikipedians do not seem very interested in sports. In the Spanish map (Figure 17), Chilean politicians are very dominant, as is the director Pedro Almodovar, and US directors and actors. Soccer, tennis and basketball players are also clearly recognisable. The French living people network (Figure 18) puts a premium on French politicians, confirming the cultural cliché of the French being focused on French affairs.

## 7 Putting it all together: wiki-evolution

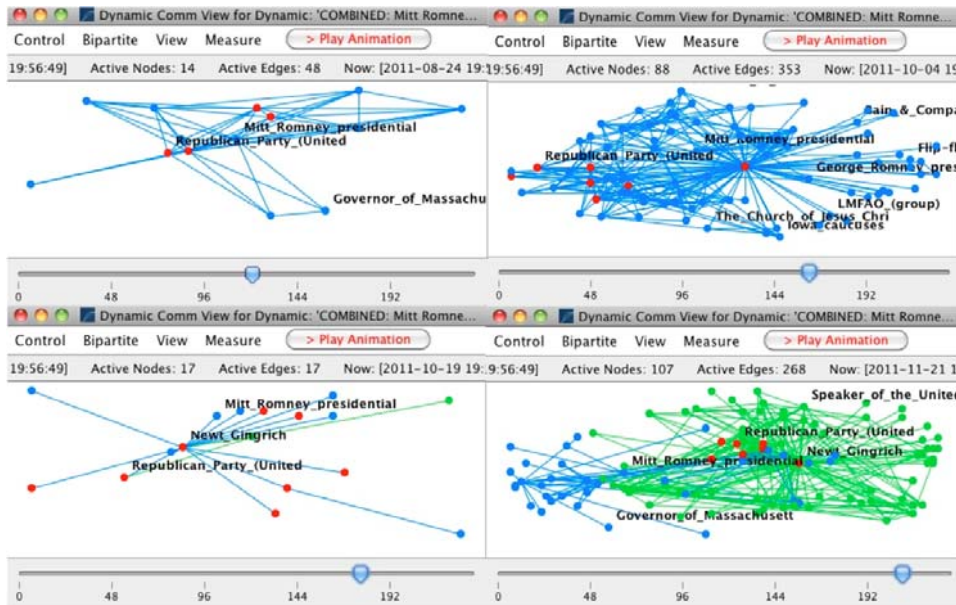
Based on the novel algorithms described in this paper, we have added a new dynamic Wikipedia mapping function to Condor (Gloor and Zhao, 2004). The new Wiki-evolution feature collects all links originating and/or pointing to one or more Wikipedia articles over a given time period.

Figure 19 and Figure 20 illustrate four months in the competition between Republican presidential candidates Mitt Romney and Newt Gingrich, through the lens of the Wikipedians. Figure 19 shows the bidirectional link network of the most important articles about the two contenders for Republican US presidential candidate. The nodes are coloured by actuality [the more edits they have in the last 14 days, the darker the red (0 to 20 edits is grey, 21 to 50 edits is orange, over 50 edits is dark red)].

**Figure 19** Bidirectional link network originating from ‘Newt Gingrich’ and ‘Mitt Romney’ Wikipedia pages (see online version for colours)



**Figure 20** Screenshots of wiki-evolution movie visualising US Republican presidential candidates Mitt Romney (blue) and Newt Gingrich (green) (see online version for colours)



As can be seen, the articles about the two candidates have attracted the most edits, but also the rock group LMFAO (whose singer SkyBlu was involved in an altercation in a plane with Romney in 2010), Pope John Paul II (who shares the distinction with Gingrich of having been chosen by Time Magazine as a Person of the Year) have had lots of edits.

Among the other presidential contenders, only Herman Cain has attracted a similar number of recent edits.

The Condor movie screen shots in Figure 20 show four months, August 2011 to December 2011, in the candidates' life reflected through Wikipedia. The pages linked to Mitt Romney are in blue, the pages linked to Newt Gingrich are in green.

The timeline comes out marvellously: It starts with the Ames Straw Poll (August 13, 2011) where both candidates did very poorly. It then shows Mitt Romney embedded into the network of other candidates, with an explosion of activity around September 18 (his shuffle with the surging Rick Perry candidacy). Discussions about Bain Capital and his membership in the Mormon Church are continuous editing topics. October 8th the Newt Gingrich candidacy takes off, slowly at first, but then really exploding October 24th, when he and Herman Cain, the favourite at that time, accepted a tea party sponsored debate. It subsequently illustrates the fight between Romney and Gingrich, taking punches at each other, also illustrating Gingrich's link as a former Speaker of the House, and Tim Pawlenty's support of Romney's campaign.

## 8 Outlook

Cultural influences are deeply engrained in each of us, influencing how we act and respond to external influences. Analysing and visualising Wikipedias from different languages and cultures opens up a unique window into different cultures and their evolution. Not only can we compare for example subtle cultural differences between countries from the same geographical region, e.g., between Korea, China and Japan, but also from vastly different influence spheres such as the Spanish or German cultural spaces, where we can trace the overlapping influences, e.g., of the emergence of democracy in Chile and Spain. We are convinced that we have just scratched the surface with WikiMaps, in obtaining a fascinating real-time window into current history and culture, spotting trends as they are unfolding right under our noses.

## References

- Chan, B., Talbot, J., Wu, L., Sakunkoo, N., Cammarano, M. and Hanrahan, P. (2009) 'Vispedia: on-demand data integration for interactive visualization and exploration', *ACM SIGMOD'09*, 29 June–2 July, Providence, Rhode Island.
- Fuehres, H., Gloor, P., Henninger, M., Kleeb, R. and Nemoto, K. (2012) 'GalaxySearch – discovering the knowledge of many by using wikipedia as a meta-search index', *Proceedings CI 2012, Collective Intelligence*, 16–18 April, Cambridge MA.
- Fruchterman, T. and Reingold, E. (1991) 'Graph drawing by force-directed placement', *Software Practice and Experience*, November, Vol. 21, No. 11, pp.1129–1164.
- Ganjisaffar, Y., Javanmardi, S. and Lopes, C. (2009) 'Review-based ranking of wikipedia articles', in *Proceedings of the 2009 International Conference on Computational Aspects of Social Networks (CASON '09)*. IEEE Computer Society, pp.98–104, Washington, DC, USA.
- Gloor, P., Laubacher, R., Zhao, Y. and Dynes, S. (2004) 'Temporal visualization and analysis of social networks', *NAACSOS Conference*, 27–29 June, Pittsburgh PA, North American Association for Computational Social and Organizational Science.
- Gloor, P. and Zhao, Y. (2004) 'TeCFlow – a temporal communication flow visualizer for social networks analysis', *ACM CSCW Workshop on Social Networks. ACM CSCW Conference*, 6 November, Chicago.

- Holloway, T., Bozicevic, M. and Börner, K. (2007) 'Analyzing and visualizing the semantic coverage of Wikipedia and its authors', *Complexity*, January–February, Vol. 12, No. 3, pp.30–40, DOI: 10.1002/cplx.20164.
- Lee, K., Kim, H., Jang, C. and Kim, H. (2008) 'Folksoviz: a subsumption-based folksonomy visualization using wikipedia texts', *World Wide Web Conference – WWW*, pp.1093–1094.
- Iba, T., Nemoto, K., Peters, B. and Gloor, P. (2009) 'Analyzing the creative editing behavior of wikipedia editors through dynamic social network analysis', *Proceedings COINs 2009, Collaborative Innovations Networks Conference*, 8–11 October, Savannah, GA.
- Nemoto, K. and Gloor, P. (2010) 'Analyzing cultural differences in collaborative innovation networks by analyzing editing behavior in different-language Wikipedias', *Proceedings COINs 2010, Collaborative Innovations Networks Conference*, 7–9 October, Savannah, GA.
- Nunes, S., Ribeiro, C. and David, G. (2008) 'WikiChanges: exposing Wikipedia revision activity', *Proceedings of the 4th International Symposium on Wikis (WikiSym '08)*, ACM, New York, NY, USA, Article 25, 4 pages.
- Shneiderman, B. and Aris, A. (2006) 'Network visualization by semantic substrates', *IEEE Transactions on Visualization and Computer Graphics*, October, Vol. 12, No. 5, pp.733–740.
- Viégas, F., Wattenberg, M. and Dave, K. (2004) 'Studying cooperation and conflict between authors with history flow visualizations', *ACM Conference on Computer-Human Interaction (CHI)*.
- Wasserman, S. and Faust, K. (1994) *Social Network Analysis*, Cambridge University Press, Cambridge, UK.
- Weld, D.S., Wu, F., Adar, E., Amershi, S., Fogarty, J., Hoffmann, R., Patel, K. and Skinner, M. (2008) 'Intelligence in Wikipedia', *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*.
- Yeh, E., Ramage, D., Manning, C., Agirre, E. and Soroa, A. (2009) 'WikiWalk: random walks on Wikipedia for semantic relatedness', *Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing (TextGraphs-4)*, Association for Computational Linguistics, Stroudsburg, PA, USA, pp.41–49.