

Wikimaps: Dynamic Maps of Knowledge

Reto Klee^{2,3}, Peter A. Gloor², Keiichi Nemoto^{1,2}

¹Fuji Xerox Co., Ltd., Tokyo, Japan

²MIT Center for Collective Intelligence, 5 Cambridge Center, Cambridge MA 02139

³Northeastern University, Boston, MA

{rkleeb, pgloor, nemoto}@mit.edu

Abstract We introduce Wikimaps, a tool to create a dynamic map of knowledge from Wikipedia contents. Wikimaps visualize the evolution of links over time between articles in different subject areas. This visualization 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, semantic network, context-analysis, graph, social network analysis.

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 analyzed collaboration among Wikipedia authors when creating new Wikipedia articles, for example studying how they collaborate as COINs in different cultures (Nemoto & 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 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.

The usage and combination of various information that is either directly available, or can be gathered through calculations based on the data that is available allows the discovery of new, previously hidden information. The visualizations built on this information will contribute greatly to 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 they are passionate about. In addition, by constructing co-authorship networks we

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

In the first phase of the project, we focus on creating dynamic visualizations of the article-link network. To achieve a reasonable animation of the data we addressed 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 visualization of these graphs over chosen time-periods.

Besides the challenges related to gathering of the data, we also work on a second research question: How do we find the most prominent, interesting and relevant pages about a chosen subject. 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. We propose an algorithm 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 network 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 visualization of information based on Wikipedia. Chan et. Al. created Vispedia (Chan et. al 2009), a mashup interface to visualize 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 delicious¹ tags. (Holloway et. al. 2007) use the Wikipedia article categories to construct a network map of all pages in Wikipedia, using color-coding to optically highlight different dimensions such as edit times, or author activity. Viegas and Wattenberg (2004) have built a well-known system to create visualizations 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.

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 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.

¹ <http://delicious.com>

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 reasonable hardware resources. The problem with these datasets is that they don't 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. The issue with these complete dumps is unfortunately their size, these 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 rather a system that gradually builds a local cache that only contains the information that is absolutely necessary. This tradeoff allowed us to build a system that minimizes the amount of required local storage while still having the possibility to benefit from a local cache for repeated queries. The fetcher minimizes the number of required calls to the Wikipedia HTTP API.

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 markup 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.

In certain cases there are multiple links from article A to a different article B, we however have decided only to count a single link, the graph in the calculation system as well as the one in the visualization have unweighed edges.

Table 1. Properties of the two Graphs that are used in Wikimaps

| Filtering-Properties | Visualization-Properties |
|--------------------------------------------|------------------------------------------------------------------------------------|
| - Simple directed graph - Uniform nodes | - Simple undirected graph - Radius of the nodes is proportional to the indegree |

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. 2011). 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 powers WikiMaps and creates the networks of article nodes is divided into two steps. 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. The upper limit of the number of nodes in the reduced graph is set according to the work of (Shneiderman & Aris 2006) to allow optimal comprehension by the users of the application.

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 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 1000 and 2500 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 to reduce the list of potential articles to the most interesting subset, combines the best candidates from three different pools. Based on well-known user interface norms (Shneiderman & 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. 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.

Visualizing the Graph

The final graph is visualized by the JavaScript/SVG based Protovis² framework. The framework simplifies the arrangement of nodes and edges according to a force-based layout (Fruchterman & 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 modern web browser to track the evolution of a given subject over the course of time. The algorithm that we chose (Fruchterman & Rheingold 1991) to arrange the nodes makes it very simple to visually recognize clusters that are related to a specific sub-topic and to discover the increase or decrease of the connectivity of said clusters.

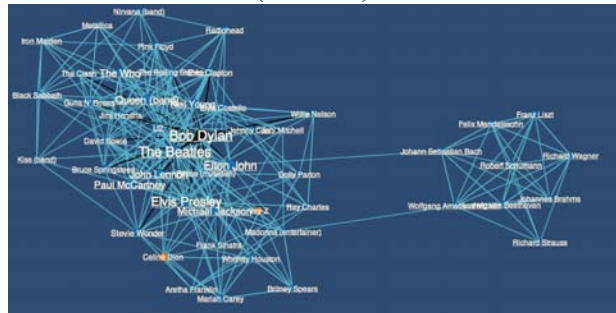
The following example displays a case of a decrease in connectivity between two subject clusters on Wikipedia. In this specific case the semantic network was built based on all musician related categories. The network clearly separates into a cluster of more modern pop and rock oriented music and second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

² <http://protovis.org>

Figure 1. Important Musicians on Wikipedia
(February 2009)



Figure 2. Important Musicians on Wikipedia
(June 2011)

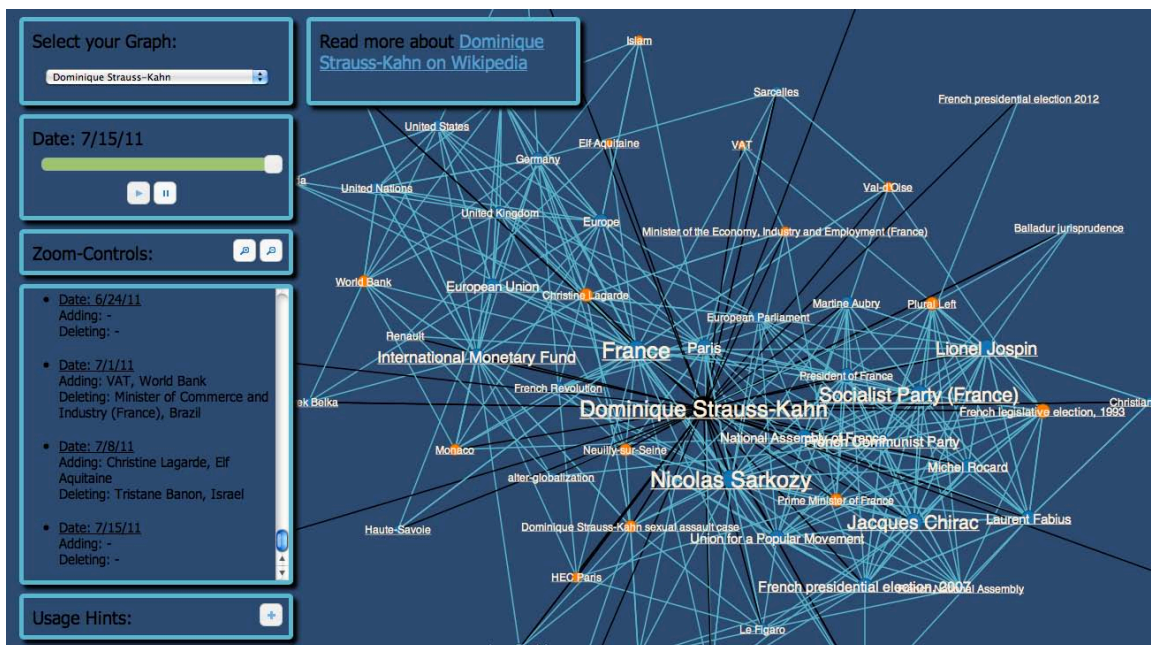
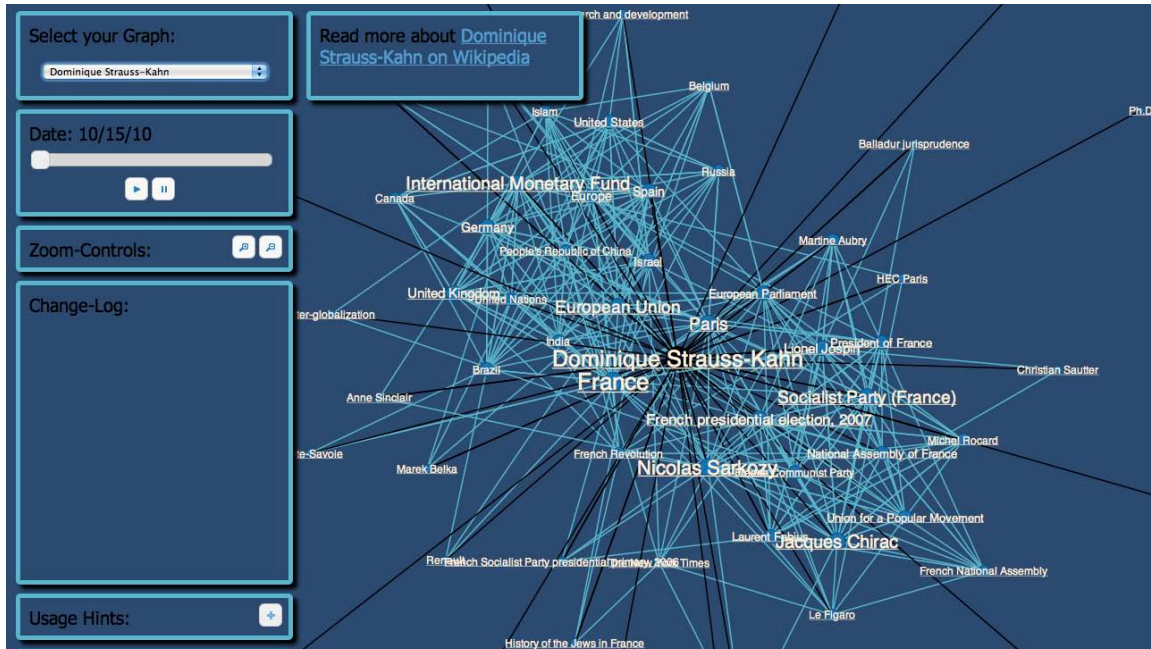


Our experimentation with the combination and the weighing of different matrices and graph properties lead to highly relevant semantic networks for the given subjects. The browser based visualization of these networks allows interested people to simply get an overview of related subjects as well as their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

The following two screenshots (figures 3 & 4) show the complete user-interface of the alpha version of the WikiMaps application. It allows the selection of a precalculated 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 (“Dominique Strauss-Kahn sexual assault case”) appears in the graph. We can further recognize 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.

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



4 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” (or 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 whole 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 10 editors based on the number of edits on this article, and then we gathered the top 10 articles of each editor. We repeated this steps recursively up to 3 degrees of separation from the start point. Restricting this analysis to a certain period of time (e.g. one month starting Jan. 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 Jan. 2010 to Nov. 2010. Each node corresponds to an article in Wikipedia. We draw an edge between articles A and B if there are at least 2 editors who made edits both on article A and article B.

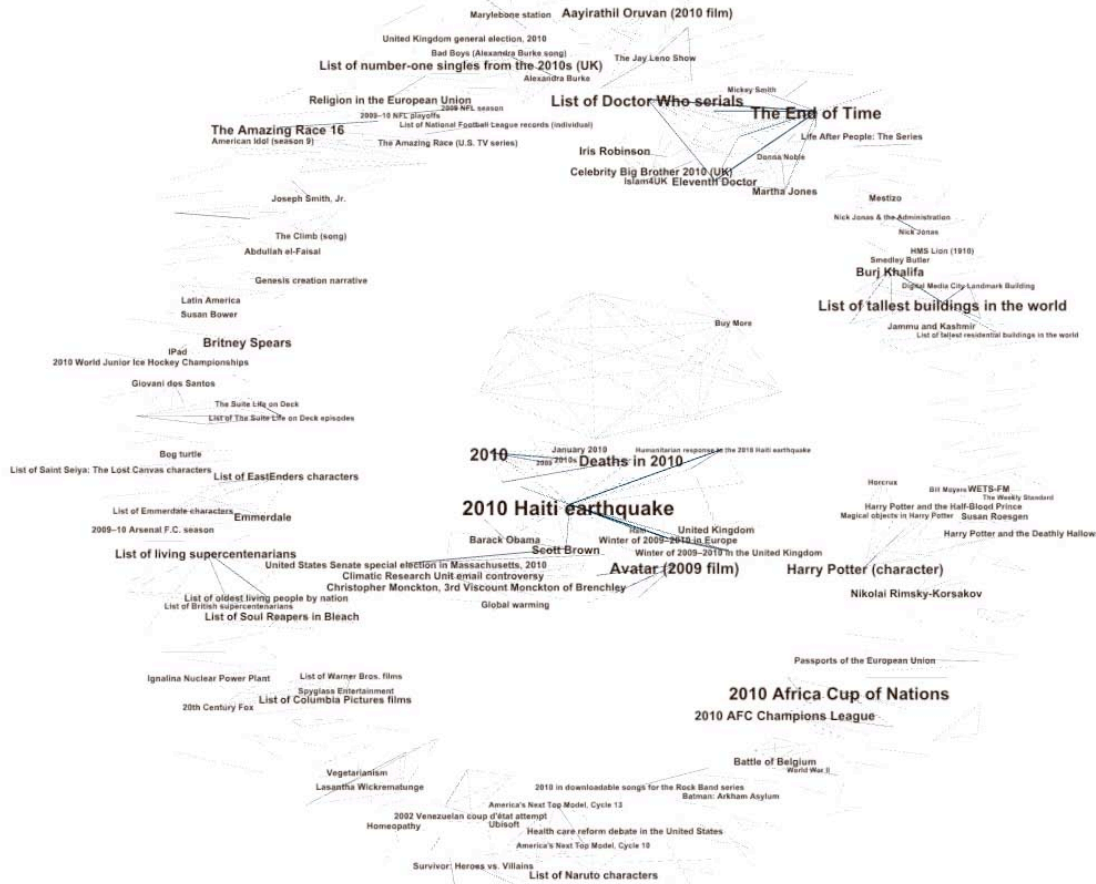


Figure 5. W5 Map for January 2010

Fig 5 and 6 show our results. Each map was drawn by Gephi, an opensource graph visualization tool (www.gephi.org). The size of the article title was determined by the undirected PageRank score of the W5 network.

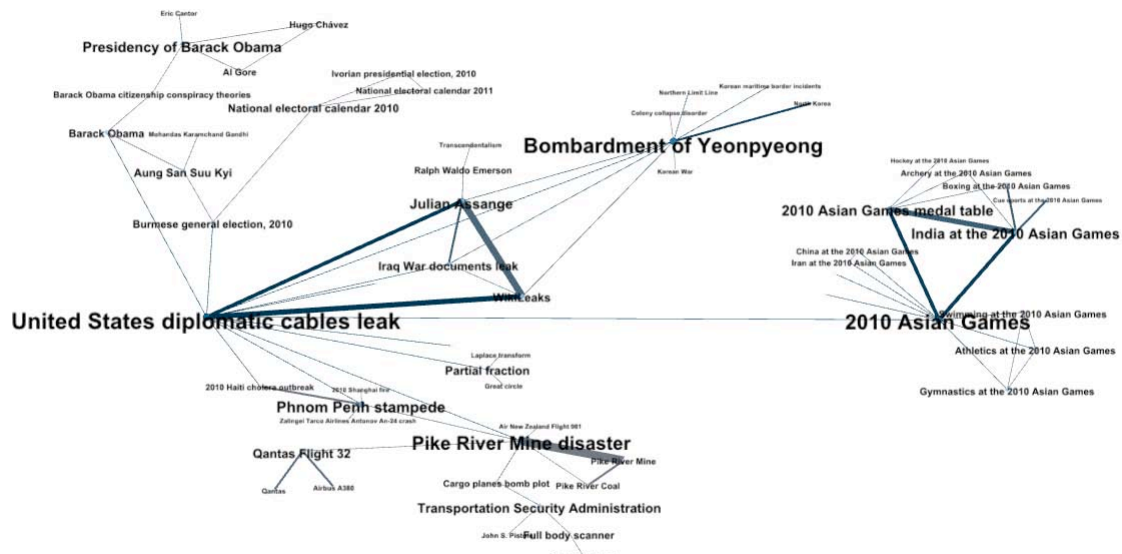


Figure 7. Zooming on on the W5 Map for November displays Wikileaks and the Bombardment of Yeonpyeong

The W5 map of Nov. 2010 (figure 7) 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 United States 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.

5 WikiPeopleMaps

Our third way of analyzing 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 478382 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 link 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 & 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 visualized, we only included articles in the visualization having more than a set 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 visualized. Therefore it varies at different points in time.

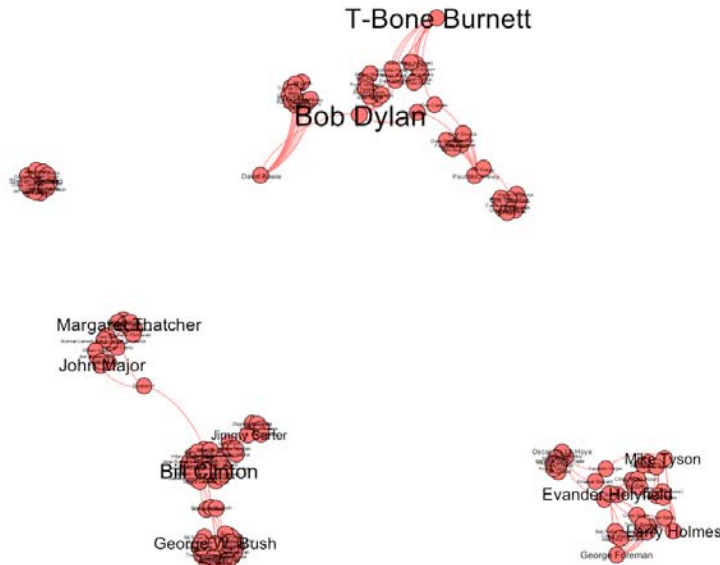


Figure 8. Living people network in the English Wikipedia in 2003 (Threshold: 10 by in-degree)

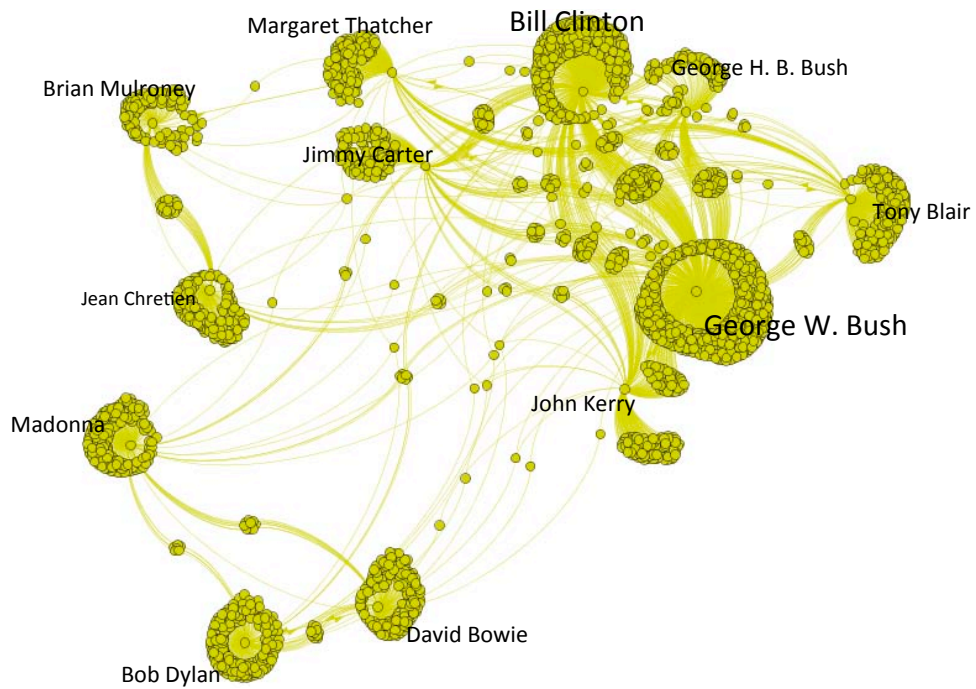


Figure 9. Living people network in the English Wikipedia in 2006 (Threshold: 150 by in-degree)

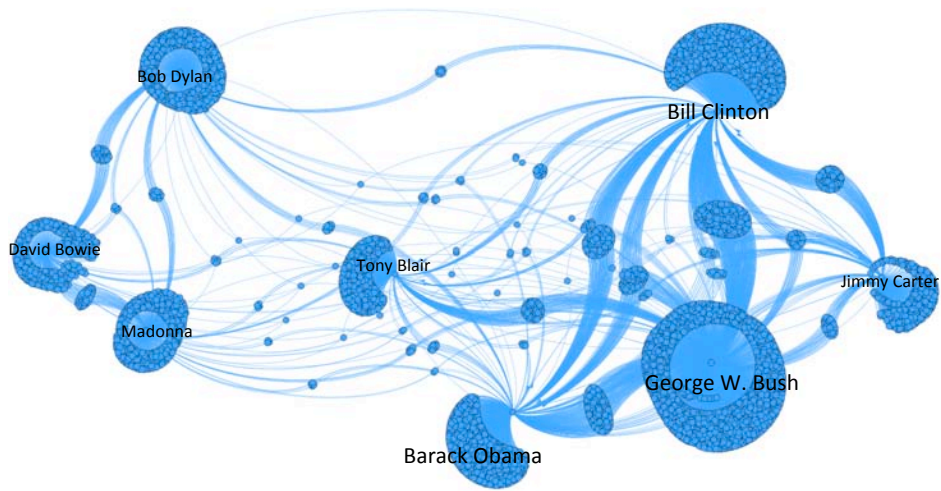


Figure 10. Living people network in the English Wikipedia in 2009 (Threshold: 500 by in-degree)

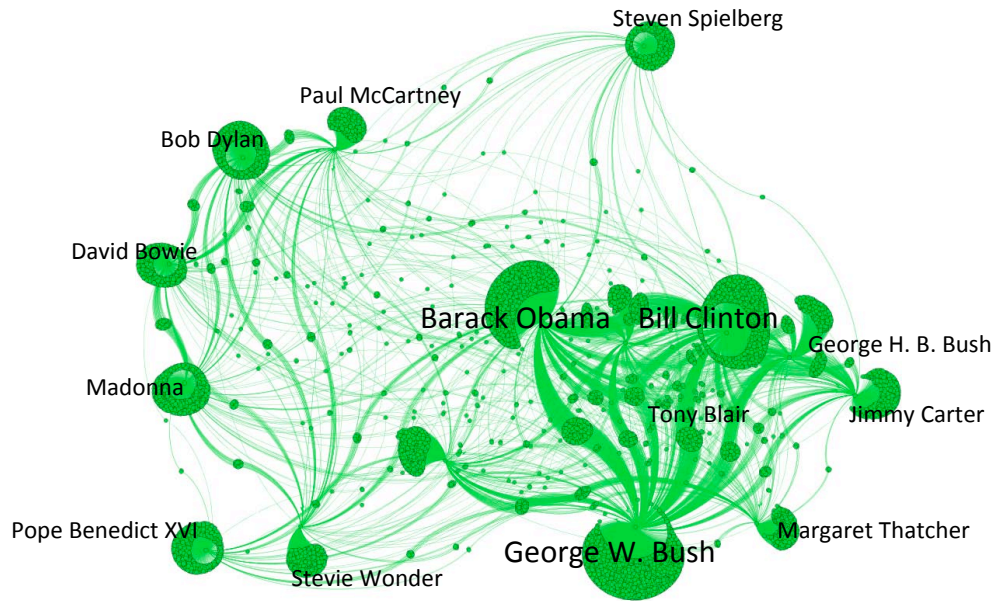


Figure 11. Living people network in the English Wikipedia in 2011 (Threshold: 500 by in-degree)

Figures 8 – 11 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.

We also visualized living people networks in the different-language Wikipedias. Table 3 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. 1900 births) and death in every year (e.g. 2000 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.

Table 3. Number of “living people pages” in the different-language Wikipedias

| Language | # of Living People articles (Feb. 2011) | # of articles (Feb. 2011) | Ratio |
|----------|-----------------------------------------|---------------------------|-------|
| 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% |

As table 3 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 egoboosting people to put up their own people page to promote themselves. We have studied this behavior in earlier work (Iba et. al. 2009).

Next we constructed the living people network for March 2011 for each of the countries listed in table 3. 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.

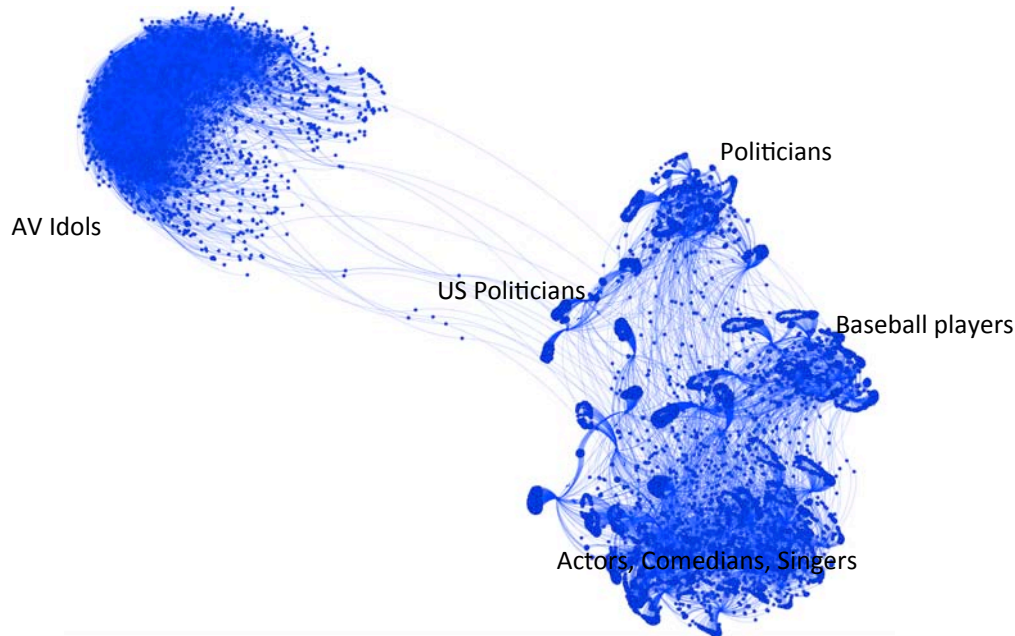


Figure 12. Japanese Living People Network in 2011 (Threshold: 150 by in-degree)

In the Japanese Wikipedia (figure 12), 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.

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 14), 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 extremely well. In the Korean Wikipedia (figure 13), 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.

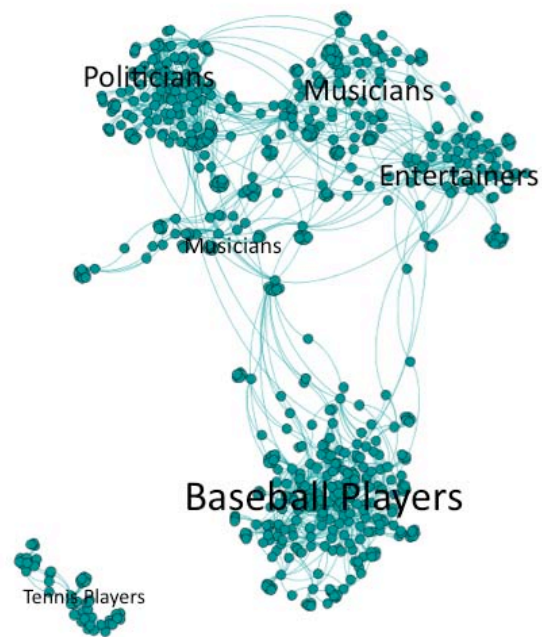


Figure 13. Korean Living People Network in 2011 (Threshold: 10 by in-degree)

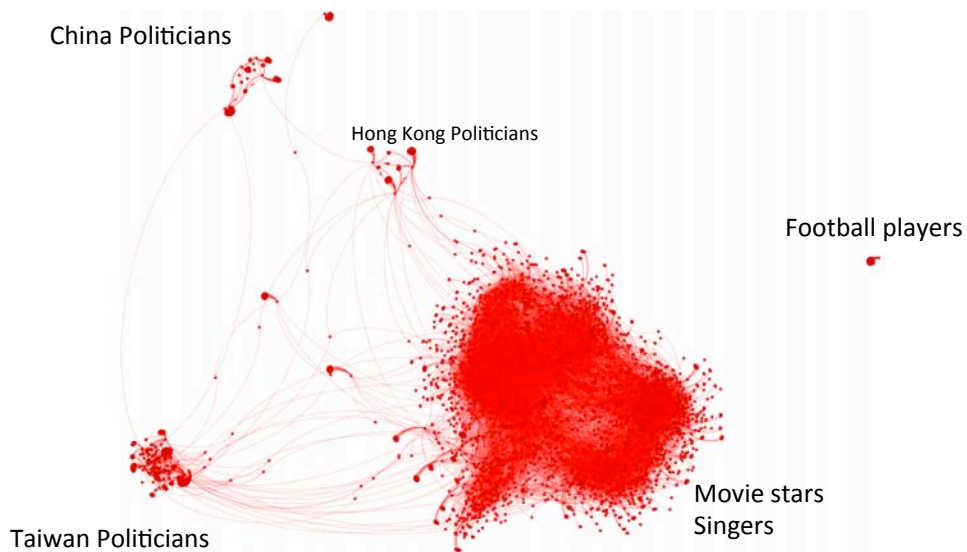


Figure 14. Chinese Living People Network in 2011 (Threshold: 30 by in-degree)

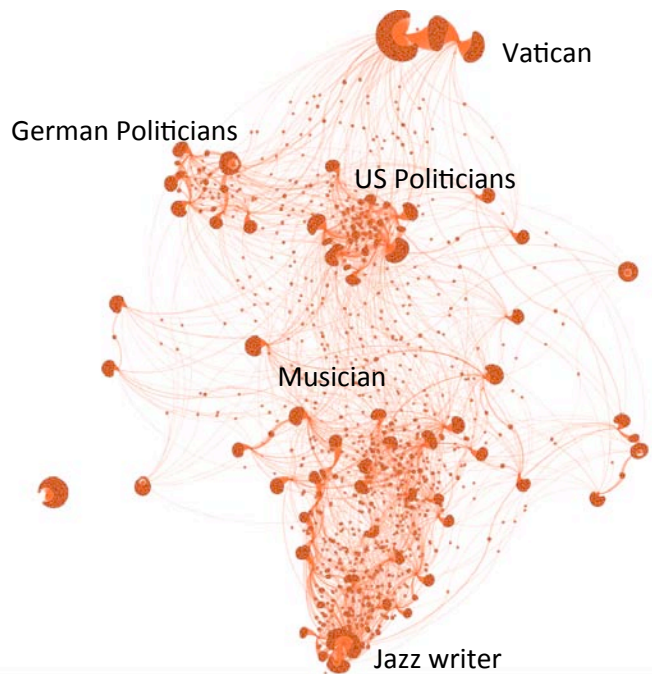


Figure 15. German Living People Network in 2011 (Threshold: 150 by in-degree)

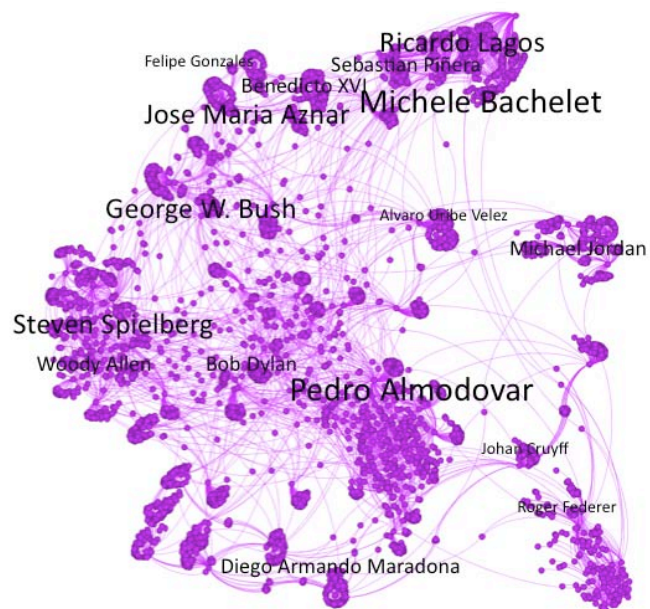


Figure 16. Spanish Living People Network in 2011 (Threshold: 50 by in-degree)

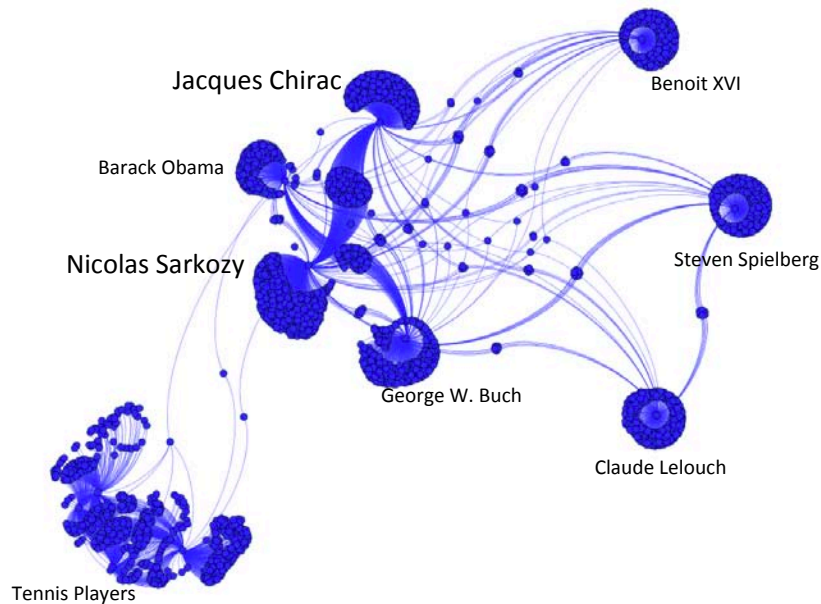


Figure 17. French Living People Network in 2011 (Threshold: 250 by in-degree)

In the German Wikipedia people map (figure 15), the top three recognizable 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 16), Chilean politicians are very dominant, as is the director Pedro Almodovar, and American directors and actors. Soccer, tennis, and basketball players are also clearly recognizable. The French living people network (figure 17) puts a premium on French politicians, confirming the cultural cliché of the French being focused on French affairs.

6 Outlook

Cultural influences are deeply engrained in each of us, influencing how we act and respond to external influences. Analyzing and visualizing 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.

7 References

- Chan, B. Talbot, J. Wu, L. Sakunkoo, N. Cammarano, M. Hanrahan, P. 2009. Vispedia: On-demand Data Integration for Interactive Visualization and Exploration. ACM SIGMOD'09, June 29–July 2, 2009, Providence, Rhode Island
- Fischbach, K. Gloor, P.A. and Schoder, D. "Analysis of Informal Communication Networks – A Case Study," Business & Information Systems Engineering, vol. 1, Apr. 2009, pp. 140- 149.
- Fuehres, H. Gloor, P. Hennigner, M. Kleeb, R. Nemoto. K. 2011. GalaxySearch – Using Wikipedia as an Index for Discovering and Visualizing the Knowledge of Many. CCI technical report.
- Fruchterman, T. Reingold, E. 1991. Graph Drawing by Force- directed Placement. SOFTWARE PRACTICE AND EXPERIENCE, Vol. 21(1 1), 1129-1164 (November)

- 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, Washington, DC, USA, 98-104.
- Holloway, T. Bozicevic, M. Börner, K. 2007. Analyzing and Visualizing the Semantic Coverage of Wikipedia and Its Authors. *Complexity*, Volume 12, Issue 3, pages 30–40, January/February 2007. DOI: 10.1002/cplx.20164
- Lee, K. Kim, H. Jang, C. 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. Gloor, P. 2009. Analyzing the Creative Editing Behavior of Wikipedia Editors Through Dynamic Social Network Analysis. *Proceedings COINs 2009, Collaborative Innovations Networks Conference*, Savannah GA, Oct 8-11.
- Nemoto, K. 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*, Savannah GA, Oct 7-9.
- Nunes, S. Ribeiro, C. and David, G. 2008. WikiChanges: exposing Wikipedia revision activity. In *Proceedings of the 4th International Symposium on Wikis (WikiSym '08)*. ACM, New York, NY, USA, Article 25 , 4 pages.
- Shneiderman, B., Aris, A. 2006. Network Visualization by Semantic Substrates. *IEEE Transactions on Visualization and Computer Graphics*, 12, 5, 733-740. (October).
- Viégas, F. Wattenberg, M. Dave, K. 2004. Studying Cooperation and Conflict between Authors with history flow Visualizations. *ACM Conference on Computer-Human Interaction (CHI)*.
- Wasserman, S., Faust, K. 1994. *Social Network Analysis*. Cambridge University Press.
- Weld, D.S. Wu, F. Adar, E. Amershi, S. Fogarty, J. Hoffmann, R. Patel, K. 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. In *Proceedings of the 2009 Workshop on Graph- based Methods for Natural Language Processing (TextGraphs-4)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 41-49.