Enhancing Web Activities with Information Visualization

Eduardo Graells-Garrido
Supervised by: Ricardo Baeza-Yates
Co-Supervised by: Mounia Lalmas
Web Research Group
Universitat Pompeu Fabra
Barcelona, Spain
eduard.graells@upf.edu

ABSTRACT

Many activities people perform on the Web are biased, including activities like reading news, searching for information and connecting with people. Sometimes these biases are inherent in social behavior (like *homophily*), and sometimes they are external as they affect the system (like *media bias*). In this thesis proposal, we describe our approach to use information visualization to enhance Web activities performed by regular people (*i.e.*, non-experts) We understand *enhancing* as reducing bias effects and generating an engaging response from users. Our methodology is based on case studies. We select a Web activity, identify the biases that affect it, and evaluate how the biases affect a population from online social networks using web mining techniques, and then, we design a visualization following an interactive and playful design approach to diminish the previously identified biases. We propose to evaluate the effect of our visualization designs in user studies by comparing them with state-of-the-art techniques considering a playful experiences framework.

Categories and Subject Descriptors

H.5.2 [**User Interfaces**]: Graphical user interfaces (GUI)

Keywords

Information Visualization; Web Activities; Web Mining; Playful Experiences; Biases.

1. THE PROBLEM

Today, users are becoming more engaged with social networks on the Web using desktops and mobile devices. However, in terms of the outcome users obtain from them, do current applications and systems actually improve users' Web activities? Several Web activities suffer from biases that diminish diversity in information, and these biases are present in the system (*i.e.*, media bias) and in users' behavior (*i.e.*, cognitive biases). Since these biases are already identified, and sometimes even quantified, in theory it could be possible to diminish their effects. In practice, it has not been easy as several actions to reduce the effects of biases can be counter-productive and, instead of enhancing Web activities, they might actually worsen them. One example of this is the *filter bubble* [28] phenomena, where recommendation systems and user-tailored information systems only show agreeable information to users. This behavior leads to group polarization and disconnection from people with opposite points of view. Direct approaches to show diverse and non-agreeable information have not worked because users do not value diversity [24]. Hence, it seems that what is displayed in systems and how it is presented is equally important. In previous work, positive effects of using information visualization techniques in these scenarios have been found [10], but those effects do not include reducing bias.

In this context, our research question is: *How to enhance Web activities using Information Visualization?* We propose to study several activities from a *web mining* perspective, to find how to introduce data-informed improvements and visualization-enabled User Interfaces (UIs). These UIs will aim at playful interaction in the context of the Web activity. We seek to generate a positive reaction in users as a way of avoiding the counter-productive reaction to potentially challenging information. This positive reaction includes change in behavior in terms of the bias we want to diminish.

The challenge that drives our research is the evaluation of the effect of visualization in a potentially challenging context. A common approach in information visualization is to design for task completion by expert users who have tool experience and understanding of the data. Having a task to perform, a new visualization design can be tested in terms of user performance (*i.e.*, accuracy, task completion time). However, in Web activities, there might not be concrete tasks to perform nor expected outcomes, *i.e.*, users may just use the Web to have a good time with friends or explore what to do on a Friday night. In general, Web activities do not have an optimal outcome, and different outcomes are not easily comparable, if any kind of measure and comparison is possible.

2. STATE OF THE ART

Currently, visualization techniques are not aimed only at expert users. Visualization has been used in journalism for decades [3], and new ways to exploit visualization in interactive narratives have been studied in *narrative visualization* [31]. Some even say the next step in visualization is *storytelling* [18]. Although not specific to visualization, *Social Data Analysis* [34] refers to the understanding of data by

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2567958.

non-experts in social settings, and visualization techniques are key for this purpose. In this aspect, *Social Data Analysis* [34] and our work are closely related to *Casual Information Visualization*, defined as *"the use of computer mediated tools to depict personally meaningful information in visual ways that support everyday users in both everyday work and non-work situations"* [30].

In previous work, algorithms and interfaces have been developed to improve access to diverse and challenging information [29, 24, 10], potentially changing user behavior [23, 10]. These approaches have been mostly direct: users are aware that the information displayed is challenging, yet still they prefer agreeable information. This direct approach has not been effective as users do not seem to value diversity [24] or do not feel satisfied with it. This behavior is explained by *cognitive dissonance* [11], a state of discomfort that affects people confronted with conflicting ideas, beliefs, values or emotional reactions.

When using interfaces that leverage information visualization, users have not improved their access to information [23], but they have improved the way they interact with others [10]. Our work takes this route following a data-informed visualization design process focused on providing positive experiences in biased scenarios.

3. PROPOSED APPROACH

We define a *Web activity* as something that might be done regularly on the Web by an individual. We focus on social, informational and leisure activities. We consider biases affecting systems (*i.e.*, *media bias*), and biases affecting users (*i.e.*, *homophily* and *selective exposure*). We focus on biases that diminish diversity and thus, we want to encourage access to more diverse sources of information.

Data-informed, Playful Visualization Design. Having a known Web activity, we identify biases that affect it based on the literature. Then, given a user population from online social networks, we perform data analysis to find how those biases are reflected on the user generated content, system provided content, and user interactions. The understanding of how users perform activities, how they interact with and within the system, and how and which biases impact the activity, allow us to define several problem-specific guidelines for our algorithms' output and visualization design. Our approach is aimed at reducing the effects of the bias, but we are not taking a direct approach as in *Casual InfoVis* [30], where data from the activity would be visualized. Instead, we use the data and its analysis as input to design a visualization that enhances the activity, possibly displaying information from other datasets or related to other activities. We embrace a playful approach by favoring it over rigid analytic visualizations to evoke positive reactions in users by giving impressions of freshness and self-organization. This is important as we focus on Web activities affected by (arguably negative) biases. To do this, we make use of two particular design approaches: *1) Organic Information Design*, a process to design visualizations of complex and dynamic information considering traits that relate to organisms like adaptation, responsiveness and movement [12] (for example, organic software visualization [26]); and *2) Designing for Serendipity Through Visualization* [32], which enumerates guidelines to create playful, interactive visualizations that encourage serendipity and entice curiosity (for example, in [9] it is used to explore faceted spaces).

4. METHODOLOGY

Our base paradigms (*organic information design* [12] and *designing for serendipity* [32]) introduce several design principles that are suitable for complex, visually rich and innovative user interfaces. Nevertheless, we strive for a *more familiar* approach because fully organic designs may seem hard to understand or too complex to be used in the context of web activities by *regular people*. As in *Casual InfoVis* [30], our designs are aimed at a wide spectrum of users, not only specialized ones nor experts. Instead of creating entirely new visualization designs, we take familiar approaches and augment them with organic features that may provoke emotional and positive responses from users. One example is the visualization in Figure 1 from one of our case studies: we have taken the familiar visualization of word-clouds (which has been on the Web since 1997 [33], and it is presumed familiar for regular users) and have added organic visual elements and playful interactive mechanisms to create a new way to interact with them.

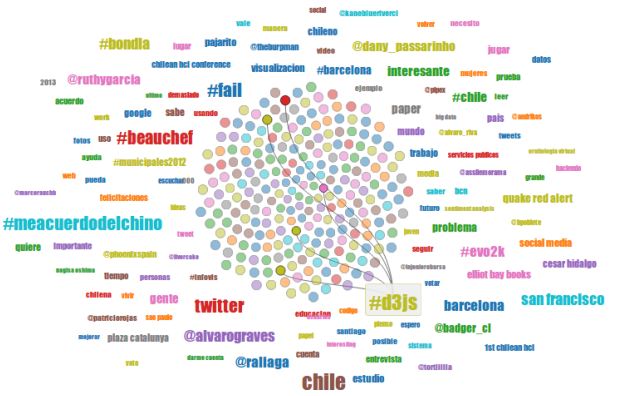

Figure 1: Data portrait [7] visualization for our case study: *Connecting with Others*. The design contains organic features and familiar elements (word-clouds). Each word is a characterizing topic for the portrayed user. Each dot represents a tweet linked to one or more topics.

Evaluation. We evaluate the quality and relevance of our algorithms' output using traditional metrics such as precision and recall [2]. We also perform user validation through crowd-sourcing. We evaluate the effect of our visualization designs in user studies: in comparison to baseline (mainstream) visualizations/user interfaces, our designs should be as effective and usable, and more playful and enjoyable. To evaluate playfulness we are considering the PLEX (*PLayful EXperiences*) framework [17, 20], which has not been used in this context before. In addition, by acknowledging that no two persons are equal, the study of *individual differences* [5] proves to be helpful, as it allows to understand which personal factors help users to enhance their web activities.

5. CASE STUDIES AND RESULTS

We started with an exploration of organic design in *Linspiration* [13], where we mixed two known visualizations for time-series: *area charts* and *spirals* [4] to create an organic

visualization. After this initial exploration, we defined case studies according to our methodology. We present three work-in-progress case studies related to web activities: *connecting with others in social networks*, *diversity-driven display of timelines*, and *exploratory search using sentiment*.

5.1 Connecting with Others

The act of connecting (or disconnecting) with others in online social networks is a web activity, as active users of social networks are regularly updating their *social graphs*. Social research has shown that, because of *homophily* and *selective exposure*, users tend to connect with similar and like-minded people. As a result, groups of users that share different points of view tend to polarize and disconnect from other groups.

Although previous efforts to use visualization have not reduced *selective exposure*, they have generated more engagement and users have been more respectful with those having opposite opinions [10]. An indirect approach to content diversity might give different results in terms of behavior and still provide positive results in terms of user engagement. Our indirect approach is based on recommending people and content about *intermediary topics* or shared interests between people of opposing views in sensitive issues. Our thinking is that if a connection based on those shared interests is established, it could be possible that the user will receive potentially challenging content in the future. This content could be better tolerated because of the *primacy effect* in impression formation [1], or, as the popular interpretation says: *first impressions matter*. In this way, we are exploiting homophily to encourage connections with dissimilar people.

Problem Statement. Our research question is: *How to encourage connection to dissimilar people in OSNs?* Our approach to answer this is outlined as follows: *1)* find what intermediary topics are feasible to recommend content and conneting two persons with opposite views on sensitive issues, *2)* define an algorithm to generate those recommendations, and *3)* visualize those recommendations in a playful way.

Data Analysis. We performed data analysis on a dataset crawled from Twitter in relation to discussions about abortion in Chile [14]. Using the vector space model [2], we define a methodology to establish an abortion stance for any Twitter account from Chile as a linear combination of two opposite stances: `#prochoice` and `#prolife`. Using topic modelling we demonstrated the existence of *intermediary topics*, *i.e.*, those with high diversity of users and high *betweenness centrality*.

Visualization Design. We visualize user profiles using *data portraits* [7], which are *"abstract representations of users' interaction history"* [35]. Our data portrait paradigm is aimed at creating a "global image" of a Twitter account, an image that depends on what identity users want to project to others. Existing work on data portraits have focused either on disruptive and artistic designs, or in solely analytic ones. We acknowledge that regular users might exhibit resistance when faced to disruptive changes in the interfaces they are used to. Therefore, as shown in Figure 1, we propose to add a new stimuli to a familiar element, *word-clouds* [33], which changes in a substantial way the interaction with it, while providing a friendly and evocative appearance based on organic patterns. By injecting recommended tweets in the visualization, we nudge users to read content from people who may have opposite views in sensitive issues, while still being relevant according to their preferences.

Pilot Study. We built a prototype content-based recommender system that ranks recommendations using a weighted combination of *topical relevance* and diversity based on opposing views on a concrete sensitive issue. We performed a pilot study were the sensitive issue was abortion and the user population on the social network and participants were from Chile. Participants were divided into three groups. The first one was a baseline, where the data portrait consisted of an interactive word-cloud and two standard list of tweets: one from the portrayed users, and other with recommendations considering topical relevance only. In the second group (*treatment I*) the recommendations considered topical relevance and stance differences. In the third group (*treatment II*) the data portrait was based on our design and the recommendations considered topical relevance and stance differences. Our results indicate that user experience with our data portrait design is comparable to a familiar baseline. The injection of content from people with opposite views had a potentially negative effect on the emotional reaction of users when comparing the baseline and the first treatment. However, our portrait design seemed to recover the emotional reaction to its previous levels. We also found hints on how individual differences are related to recommendation attributes: *openness* is linked to perceived interestingness, and engagement with our design is related to perceived serendipity.

Future Work. Next steps include formalizing our recommender system and perform crowd-sourced evaluations of the recommendations. Then, we will create an open-system for Chilean users where anyone will be able to join and make use of our visual design, and we will log interactions between users. We will also perform a longitudinal study to measure the effect of our visualization design and to validate the results of our pilot study.

5.2 Diversity-driven Display of Timelines

In the Twitter microblogging platform, users follow other accounts that publish micro-posts with news, opinions, and general content, among interactions between users. These micro-posts or tweets are displayed in reverse-chronologically ordered timelines. In countries with severely imbalanced distributions of population, it is likely that most populated places generate more tweets, making tweets from less populated locations to become lost in the timeline, losing the potential content richness derived from geographical diversity. Previous work has tackled the problem of generating diversity-driven timelines [6, 25], but the question of how to encourage reading diverse content has not been solved yet.

Problem Statement. Our research question is: *How can we use visualization to encourage access to geographically diverse content in user timelines?* Our outline to answer this is as follows: *1)* define a text classifier that, given a *tweet*, predicts the location its content is related to, *2)* define an algorithm to filter a timeline and return a filtered one that maximizes geographical diversity, *3)* design a visualization to encourage reading of more diverse content in timelines by displaying the filtered timeline.

Data Analysis. To build a geographically diverse timeline we need a way to identify where a tweet comes from and what is the location related to its content (both are not necessarily the same). Although tweets can include geographic infor-

mation, the portion of tweets with geographic coordinates is very low. In [15] we studied how tweets are distributed in Chile, and found that only 7% of tweets generated from Chilean accounts include geographical coordinates. A trivial classifier that associates all tweets to Santiago (the capital) already obtains an accuracy of 58% at city level. We defined a model to normalize social media information according to geography, by using the vector space model with TF-IDF weighting [2] to model locally characteristic content in several locations. Using this model we were able to: *1)* predict the location related to a tweet with better performance than the non-diverse baseline, and *2)* determine how the physical population imbalance is reflected on the virtual population of Twitter.

Current Status and Future Work. In addition to improving the classifier built in [15], we will define an algorithm to generate diverse timelines, maximizing geographical diversity according to *Shannon entropy* [16]. Then, we will design a playful visualization to explore content generated on the social network using a geographical facet. Given that the virtual population is imbalanced, the challenge we face is to encourage browsing of information related to less populated locations. This is not an easy task as the core user experience in timeline browsing is focused at recency and not geography. We will address this trade-off by providing a playful visualization for faceted exploration that emphasizes content from less populated locations.

5.3 Exploratory Search Using Sentiment

Figure 2: Search interface from the *Exploratory Search using Sentiment* case study. We used parallel coordinates to display *ambivalence* in search results.

Even though search engines have been present for many years on the Web, today most of them still have the initial text-based interface, in spite of the emergence of several paradigms in information seeking. One of those paradigms is *Exploratory Search* [21], where a concrete information need is not always present. One way to support exploratory search is by using faceted search interfaces, where information seekers have access to several orthogonal dimensions of the information space even when there is no explicit information need. However, its implementation requires a structure in the underlying data that is not always available. A solution is to extract meta-data from the information space to provide the needed structure. We adopt this approach to build a facet for an unstructured information space, by using attributes annotated in text documents calculated through *sentiment analysis* [27].

We consider exploratory search using sentiment analysis in the context of social networks based on user reviews. In these networks, users write reviews and give scores to products, services and places in a city. The quantity of reviews and its scores influence future consumption and check-ins by users. In particular, users tend to give more positive reviews and to go to places that already have many positive reviews. This is known as the *bandwagon effect* [19]. In this case study, we propose to use visualization of sentiment-facets to encourage exploration of places that might be good even with a small number of reviews but are ignored because of the *bandwagon effect*.

Problem Statement. Our research question is: *How to reduce the bandwagon effect by using visualization in a faceted exploratory search context?* We chose sentiment as primary facet because reviews are inherently based on them. However, most sentiment and score depictions focus on simple visualizations and discrete scores from 1 to 5. Such depictions do not consider the *ambivalence* present in text, which means that a document may have both positive and negative content at the same time. Considering this scenario, the outline of this case study is as follows: *1)* we make use of ambivalence to normalize review scores and text into a structured dataset in terms of sentiment, *2)* we build playful *visualization widgets* [8] for a sentiment facet in the context of exploratory search.

Formative Study and Future Work. To start to define an approach to answer our research question we performed a formative user study. In a sentiment-based exploratory search context we tested two known visualizations: *scatter plots* and *parallel coordinates* against a text-based baseline (see Figure 2). As information space for our pilot study we chose Wikipedia, an open encyclopedia where anyone can contribute and edit articles. Wikipedia is a prominent social media platform, which contains articles with inherent sentimental content [22]. Our qualitative results show that users are favorable to the proposed visual approaches in comparison to a text-based baseline. Our quantitative results show that, when using our system, people who likes to explore lose track of time in a positive sense, whereas people focused on completing the tasks at hand over-estimate how much time they will spend on it. Our results suggest that text filters and scatter plots are more suitable for people focused on exploration whereas parallel coordinates are more beneficial to people focused on task completion. Therefore, we will base our ambivalence visualization widgets in scatter plots, with added organic visual features and playful interaction mechanisms. However, before designing the final visualization, we will propose an algorithm to normalize review scores and text from a user generated review dataset.

6. CONCLUSIONS AND FUTURE WORK

In this thesis proposal we address the problem of enhancing web activities with Information Visualization. The core idea behind our proposal is that playful visualization has the potential to evoke positive reactions in users, increasing engagement in a way that allows us to inject interaction mechanisms that, at first, may seem counter-productive, like recommending people with opposite views on sensitive issues.

We presented on-going results of three case studies, with encouraging preliminary results. Our web mining approach to analyze the user population and its generated content has delivered results that validate the assumptions stated in

our motivations (*i.e.*, *intermediary topics* and *virtual population reflecting the imbalance of the physical population*). However, the main contribution of this thesis proposal is the design approach and its evaluation from a playful experiences framework, which remains to be tested. Although we have embraced disruptive and complex design frameworks, our approach on familiar visualization elements with added organic and playful elements has obtained good results with non-experts users, including links between individual differences and perception of how engaging and useful our user interfaces are. The significance of these results will be tested in larger scale user studies to be performed on open-systems.

Acknowledgments. This work was partially funded by Grant TIN2009-14560-C03-01 of the Ministry of Science and Innovation of Spain.

7. REFERENCES

[1] Solomon E Asch. Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, 41(3):258, 1946.

[2] R. Baeza-Yates and B. Ribeiro-Neto. *Modern information retrieval: the concepts and technology behind search, 2nd. Edition*. Addison-Wesley, Pearson, 2011.

[3] Alberto Cairo. *The Functional Art: An introduction to information graphics and visualization*. New Riders, 2012.

[4] John V Carlis and Joseph A Konstan. Interactive visualization of serial periodic data. In *Proceedings of the 11th annual ACM symposium on User interface software and technology*, pages 29–38. ACM, 1998.

[5] Chaomei Chen, Mary Czerwinski, and Robert Macredie. Individual differences in virtual environments—introduction and overview. *Journal of the American Society for Information Science*, 51(6):499–507, 2000.

[6] Munmun De Choudhury, Scott Counts, and Mary Czerwinski. Identifying relevant social media content: leveraging information diversity and user cognition. In *Proceedings of the 22nd ACM conference on Hypertext and hypermedia*, pages 161–170. ACM, 2011.

[7] Judith Donath, Alex Dragulescu, Aaron Zinman, Fernanda Viégas, and Rebecca Xiong. Data portraits. In *ACM SIGGRAPH 2010 Art Gallery*, pages 375–383. ACM, 2010.

[8] M. Dörk, S. Carpendale, C. Collins, and C. Williamson. Visgets: Coordinated visualizations for web-based information exploration and discovery. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1205–1212, 2008.

[9] Marian Dörk, Nathalie Henry Riche, Gonzalo Ramos, and Susan Dumais. Pivotpaths: Strolling through faceted information spaces. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2709–2718, 2012.

[10] Siamak Faridani, Ephrat Bitton, Kimiko Ryokai, and Ken Goldberg. Opinion space: a scalable tool for browsing online comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1175–1184. ACM, 2010.

[11] Leon Festinger. *A theory of cognitive dissonance*, volume 2. Stanford university press, 1962.

[12] Benjamin Jotham Fry. *Organic information design*. Massachusetts Institute of Technology, 2000.

[13] Eduardo Graells and Alejandro Jaimes. Lin-spiration: using a mixture of spiral and linear visualization layouts to explore time series. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, pages 237–240. ACM, 2012.

[14] Eduardo Graells-Garrido, Mounia Lalmas, and Daniele Quercia. People with opposing views can share common interests. In *Proceedings of the 23rd international conference companion on World Wide Web*. ACM, 2014. To appear.

[15] Eduardo Graells-Garrido and Bárbara Poblete. #santiago is not #chile, or is it?: a model to normalize social media impact. In *Proceedings of the 2013 Chilean Conference on Human-Computer Interaction*, pages 110–115. ACM, 2013.

[16] Lou Jost. Entropy and diversity. *Oikos*, 113(2):363–375, 2006.

[17] Hannu Korhonen, Markus Montola, and Juha Arrasvuori. Understanding playful user experience through digital games. In *International Conference on Designing Pleasurable Products and Interfaces*, pages 274–285, 2009.

[18] Robert Kosara and Jock Mackinlay. Storytelling: The next step for visualization. *Computer*, 46(5):44–50, 2013.

[19] Harvey Leibenstein. Bandwagon, snob, and veblen effects in the theory of consumers' demand. *The Quarterly Journal of Economics*, 64(2):183–207, 1950.

[20] Andrés Lucero, Jussi Holopainen, Elina Ollila, Riku Suomela, and Evangelos Karapanos. The playful experiences (plex) framework as a guide for expert evaluation. In *Proceedings of the 6th International Conference on Designing Pleasurable Products and Interfaces*, DPPI '13, pages 221–230, New York, NY, USA, 2013. ACM.

[21] G. Marchionini. Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4):41–46, 2006.

[22] Yelena Mejova, Ilaria Bordino, Mounia Lalmas, and Aristides Gionis. Searching for interestingness in wikipedia and yahoo! answers. *International World-Wide Web Conference (WWW)*, 2013.

[23] Sean A Munson, Stephanie Y Lee, and Paul Resnick. Encouraging reading of diverse political viewpoints with a browser widget. *ICWSM*, 2013.

[24] Sean A Munson and Paul Resnick. Presenting diverse political opinions: how and how much. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1457–1466. ACM, 2010.

[25] Sean A Munson, Daniel Xiaodan Zhou, and Paul Resnick. Sidelines: An algorithm for increasing diversity in news and opinion aggregators. In *ICWSM*, 2009.

[26] Michael Ogawa and Kwan-Liu Ma. code_swarm: A design study in organic software visualization. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1097–1104, 2009.

[27] B. Pang and L. Lee. *Opinion mining and sentiment analysis*. Now Pub, 2008.

[28] Eli Pariser. *The filter bubble: What the Internet is hiding from you*. Penguin UK, 2011.

[29] Souneil Park, Seungwoo Kang, Sangyoung Chung, and Junehwa Song. Newscube: delivering multiple aspects of news to mitigate media bias. In *Proceedings of the 27th international conference on Human factors in computing systems*, pages 443–452. ACM, 2009.

[30] Zachary Pousman, John T Stasko, and Michael Mateas. Casual information visualization: Depictions of data in everyday life. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1145–1152, 2007.

[31] Edward Segel and Jeffrey Heer. Narrative visualization: Telling stories with data. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1139–1148, 2010.

[32] Alice Thudt, Uta Hinrichs, and Sheelagh Carpendale. The bohemian bookshelf: supporting serendipitous book discoveries through information visualization. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pages 1461–1470. ACM, 2012.

[33] Fernanda B Viégas and Martin Wattenberg. Tag clouds and the case for vernacular visualization. *interactions*, 15(4):49–52, 2008.

[34] Martin Wattenberg and Jesse Kriss. Designing for social data analysis. *IEEE Transactions on Visualization and Computer Graphics*, 12(4):549–557, 2006.

[35] Rebecca Xiong and Judith Donath. Peoplegarden: creating data portraits for users. In *Proceedings of the 12th annual ACM symposium on User interface software and technology*, pages 37–44. ACM, 1999.