Emotion aware clustering analysis as a tool for Web 2.0 communities detection: Implications for curriculum development

Athena Vakali
Informatics Department
Aristotle University, Greece
avakali@csd.auth.gr

Konstantinos Kafetsios
Department of Psychology
University of Crete, Greece
k.kafetsios@psy.soc.uoc.gr

Abstract

The emergent Web 2.0 reality has advanced a new role for Web users since they now approach information in a dynamic way regulating content, opinions, and policies. Revealing, analyzing and exploiting non-evident (often hidden) communities formulated in Web social networks is crucial, since communities influence content distribution and drive Web trends and events. It is now important to overcome typical single- criterion community detection methodologies (usually originating from graph mining), and within multidisciplinary efforts advance novel multiple criteria approaches which will identify communities of high coherence and homogeneity.

In constructing such Web community indices (both now and in the future Web context) it is vital to consider human behavioral and cognitive criteria, since, it is those that affect users' activities, preferences and social interactions on the Web. We therefore argue, that within typical processing criteria (such as frequency of access, user profiling, and content semantics), we need to incorporate affective criteria which are closely connected to users' actions and social interactions. In this paper we present an emotion aware clustering approach that incorporates affect as a central component. This approach can be applied to a range of activities such as: highlighting non-obvious and evolving phenomena on the Web, improving data accessing performance, assisting the design of novel content promotion strategies, and developing targeted actions of personalized recommendation. The report identifies the scientific and technical background needed for such multidisciplinary approach on the Web 2.0 and highlights the major topics required for a competitive Web science curriculum.

1. Introduction

Deriving user communities on the Web can be beneficial due to the fact that, once detecting a user's community, several applications (such as recommendation, personalization, content outsourcing, crawl strategies, marketing, etc.) may become better targeted and focused. These applications can also exploit Web users' behavior analysis towards improving accessing performance, quality of services, users satisfaction etc.

The role of Web users has shifted from passive navigators to content regulators and moderators since Web 2.0 incorporates a wide set of tools and technologies that focus on information sharing, online collaboration and communication. Consequently, there is a significant increase in user participation in numerous established and emerging communities that now involve millions of users worldwide in online social interactions (within seminal applications like Twitter, Facebook etc). In such an emerging scene, two major community types exist: the *explicitly-defined* (i.e., obvious groups of Web users sharing a common interest such as the groups in Yahoo, Drupal , Google, LinkedIn, Facebook etc) and the *implicitly-defined* (i.e. non-obvious, often hidden and unexpected groups emerging from user interactions such as in the case of Flickr clusters involving highly correlated members attracted over a topic or an event). Explicitly-defined communities are coordinated and governed by an authoritative subject or process (such as a group inventor). The noncoordinated but self-organized implicitly-defined communities offer the ground for researching dynamic community development since they emerge from Web usage patterns and social interactions.

This report emphasizes the role of human-side criteria in community formation (and analysis) since users' decision making is based on global scale opinions which span from the individual (e.g., a circle of friends) to the universal (e.g. business surveys, focus groups, etc). Word-of-mouth on the Web is a current reality due to the emergent and large-scale user-generated media on which people express opinions (reviews, forums, discussion groups, blogs etc). In this context, analysis of emotion information is necessary since affect is the very natural human expression and a behavioral characteristic of communities and social interaction off-line. Importantly, as a recent review suggests [Derks08], computer-mediated communication isequally, emotion ridden Therefore analyses of affect-related information detection is seminal to understanding attitude and opinion convergence, in a set of documents or user-generated content and consequently, community formation.

The identified developments have important implications for developing novel, multidisciplinary, curricular structures for Web scientists. Tomorrow's educational reality calls for Web scientists who can utilize diverse knowledge bases from the information and behavioral sciences in order to anlalyze communities and be able to advance suggestions, policies and regulations. Therefore, a cross-disciplinary curriculum should include, eclectically, topics which span from computer science to the behavioral (e.g., psychology, cognitive science) and the social sciences, but this range should be crossed with joint themes which will facilitate this crossing. We claim that affect-aware community detection can be used as vehicle for supporting cross-disciplinarity and we suggest their relevant foremost topics.

2. Detecting Web 2.0 communities : Current status

Over the last decade the typical community detection and definition varied over different aspects such as popularity, semantics, and structure [Fortunato08]. In the emerging and active Web 2.0 context, communities fall within one of three types: user communities (groups of people that upload/tag/visit/comment on related resources or use related tags), tag communities (groups of tags that are semantically close or, often co-occur in resources description) and resource communities (groups of resources such as text, images, video etc that are highly related through any of social interaction patterns) [Vakali08], [Papadopoulos11].

Communities are typically defined by graphs with a set of nodes having common properties (e.g., web pages with similar topics), whereas in the context of the social Web, communities often refer to sets of closely interrelated users, resources and/or tags [Papadopoulos10]. Three crucial tasks determine Web 2.0 community detection effectiveness :

i. the definition of metrics for community members *relationships assessment is necessary* since, in Web2.0 and social networks more broadly, explicit connections may exist among users (e.g. friendships in Facebook), but implicit relationships can be exported by users' activities [Cattuto08]. Relationship similarity measures include: (i) *co-occurrence*, used on its own or in conjunction with a semantic similarity measure [Giannakidou08], (ii) *cosine similarity*, used to detect communities between resources and users with similar interests [Shepitsen08], (iii) *tf- idf*, used in methods identifying communities of blog posts [Brooks06], and (iv) *hybrid* measures combining non-textual information (e.g. visual features) in a social tagging system [Li09].
ii. the introduction of *algorithmic approaches designed exclusively for communities' identification in complex networks* (such as the social Web), since it was realized that typical graph partitioning and node clustering methods cannot be practically and efficiently applied, due to their imposed restrictions (number of groups pre-specification, etc). Most significant approaches include: (i) *divisive algorithms*, that repetitively remove edges connecting nodes of different communities on the basis of some metric

[Radicchi04], and (ii) *spectral algorithms*, which exploit the algebraic properties of matrices derived from graphs and cluster nodes based on the similarity of their eigenvectors [White05].

iii. the definition of indicator measures for the evaluation of the derived communities is usually accomplished based on indicators, such as: (i) *betweenness centrality*, which evaluates quantitatively the importance of a graph's edge, (ii) *modularity*, whose maximization indicates a structure of "strong" communities, and (iii) *cut-size*, which represents the sum of weights of the edges connecting two sets of nodes [Newman06].

In summary, no global methodology for community detection supports appropriate inclusion and integration of both computational and processing with human behavior parameters.

3. Emotion-aware Web 2.0 communities detection

Web 2.0 communities should capture people's tasks relationships and interactions and therefore human behavior and emotion are paramount for such a detection process. With the advent of Web social applications, millions of people broadcast their thoughts and opinions on a variety of topics and, with this being an increasingly popular way of online communication, users heavily interact and broadcast their personal thoughts. Therefore, these applications contain highly opinionated personal commentary and the new social media offer a unique look into people's emotion-laden reactions and attitudes and how those could relate to implicit or explicit web-communities.

We argue that the emotional experience in web-communities is an important and yet untapped component of online social interaction since typically, related work on the topic has applied many variations of sentiment analysis [Liu10]. A recent example [Bolen11] has proposed performing a sentiment analysis of tweets using an extended version of a well established psychometric instrument, the Profile of Mood States (POMS). In this work, a six-dimensional vector representing the tweet's mood and aggregate mood components is extracted on a daily scale comparing results to the timeline of cultural, social, economic, and political events that took place in that defined period of time.

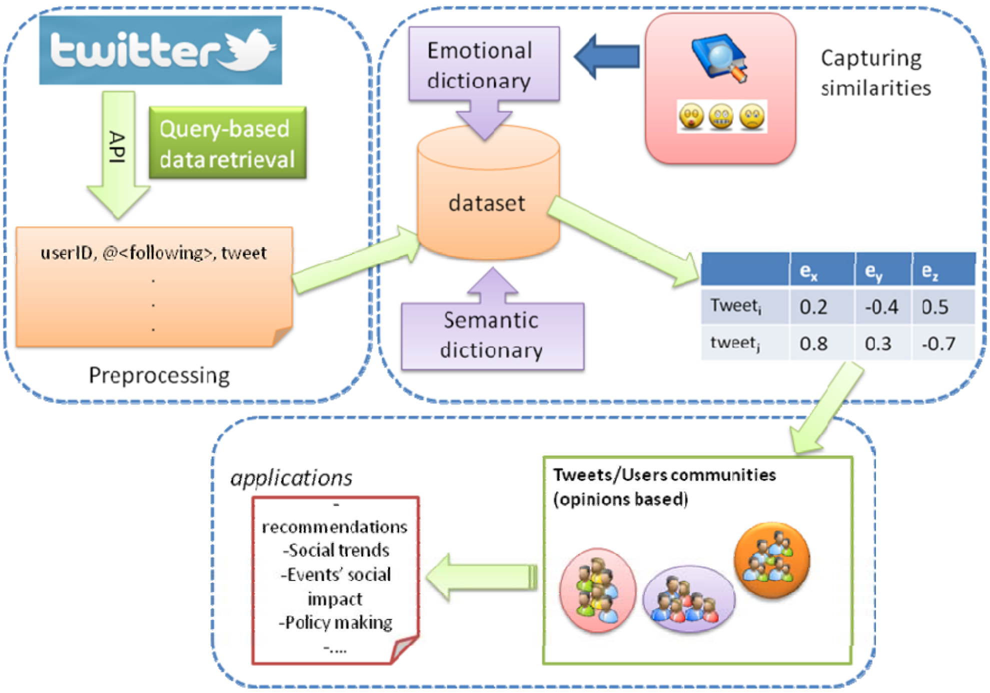

Figure 1 : An emotion-aware community detection framework

Building on analyses of the emotion characteristics of opinions and commentaries we [Tsagkalidou11] have recently proposed and tested an emotional aware clustering approach. The approach stresses the affective characteristics of online written communications as expressed in the popular Twitter microblogging application. Figure 1 summarizes our approach which focuses on the primary emotion dimensions of overarching positive or negative affect based on the semantic analysis of acceptance, fear, anger, joy, anticipation, sadness, disgust and surprise and their synonymous adjectives. The assumption is that some basic emotional words which constitute the emotions representatives can determine the sentiment (or, at least some indication) of a sentence by analyzing sentence's words similarities with emotion representatives. This emotional aware clustering which performs tweets sentiment analysis can be used to reveal communities of users and their tweets according to the degree of users' emotion expresiveness. Experimental evaluations on datasets derived from Twitter prove the efficiency of the proposed approach.

Our current, cross-disciplinary efforts centre on *affect-aware community detection*, since affect, the overarching positive or negative evaluation of a stimulus [Cacioppo99], is one of the most important dimensions of social interaction [Darwin65]; [Forgas01]. Affect is distinguished from discrete emotions in that discrete emotions concern affective reactions in relation to one's goals [Frijda86], whereas affect refers to an overarching positive or negative valence of one's feelings. Groups (and consequently communities) can be distinguished in terms of prevalent affective [Kelly01] and emotion [Seger09] information. Applying the concept of 'herding' [Raafat09] we maintain that analysis of affective information for web communities can be important in detecting of such 'herds'/communities. Affective information can be an important social cue of web communities that allows the group to exchibit 'higher order computational capacities' [Couzin07].

4. Web community detection advances have implications for Web science curriculum

The main strategic objective is to establish a unified understanding and provide a common vision for the formulation of a Web science curriculum which will encompass and bridge views from the necessary disciplines and orientations, such that topics like seminal community detection will be qualitatively supported.

Importantly, within the noted paradigm change in affective computing and the utilization of emotion aware clustering for community detection outlined in the previous section, the new web science curriculum should include a number of modules from the social and behavioural sciences (psychology, social psychology, cognitive science, statistics). In particular, implementation of current data analysis (such as the affect aware clustering) requires that Web Scientists are knowledgeable and informed on:

- social networks and the related mathematical and statistical principles;
- the social psychology of Emotion and Affect in particular as it is applied in dyadic social interaction and in groups;
- group behavior and attitude formation;
- the cognitive dynamics of interpersonal interaction

We propose an outline for an appropriately designed curriculum which will complement expertise and be driven by twofold views offered by the computer and the psychology sciences. Figure 2 encapsulates this duality and it covers the suggested here topics and joint themes as summarized next :

- **Data structures and measures involve** new Web2.0-tailored data representation structures and measures capturing users' behavior and preferences. Data representation models and structures should be based on: i) *connection subgraphs* that encapsulate interactions and relationships in a social network setting, and ii) *evolving graphs* that record unfolding usage of social and impersonal interactions over time. New measures should be highlighted to integrate usage, semantics and affect information as derived from users' interactions and groups behavior.

- **methodologies and techniques include** community detection algorithms based on newly-defined measures which integrate usage, semantics and users emotional and behavioral information towards enriching and improving communities' quality and coherence. New algorithms and techniques should cover issues raised in overlapping communities which embed affect and emotion temporal evolution.
- **Experimentation and evaluation unify** tools incorporating customized methods for data collection and preprocessing under parameterized processes which collect data from various Web2.0 sources in an affect -aware manner. Benchmarks are needed efficient testing and evaluation with scalability and behavioral modeling capabilities. Assessment and evaluation measures should enhance their functionality by affect-wise factors.
- **Technologies and applications tend** to focused social interactions capturing, implementation of human-centric recommendation engines and skills should be developed such that both affect-wise and social-aware technologies deliver high quality services and functionality.

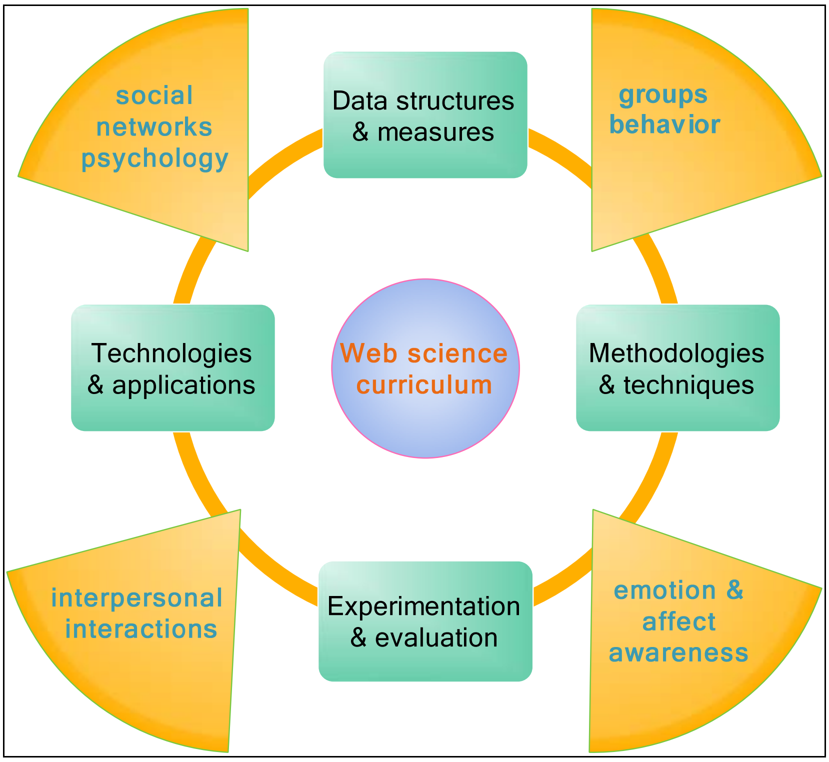

Figure 2 : A Web science curriculum outline

The above curriculum suggestions and their inherent challenges, will equip future Web science researchers and practitioners with the appropriate skills and cross-discipline expertise. It is foreseen that several sectors and markets could offer a potential ground since they are characterized by both social and computational needs as highlighted next :

- education : learning process is of collaborative nature;
- customers decisions : need peer opinions to make purchases or choices;
- business providers: need customers' opinions to improve product and need to track opinions to make marketing decisions;
- social researchers: yearn for people's reactions about social events;
- government: needs people's reactions awareness to proceed to new policies.

These emerging markets need properly educated Web scientists who will have a cross-discipline expertise and will have a firm background in topics spanning from Web data mining to affect analysis and group behavioral awareness. This report aims at highlighting the major topics needed in a Web scientist background when having affect-aware community detection and identification as a vehicle for such a challenging bridging between disciplines.

References

[Bollen11] Bollen, J., Pepe, A., Mao, H.: Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In: ICWSM 2011, arXiv: 0911.1583 (2011)

[Brooks08] Brooks, C. & Montanez, N. : Improved Annotation of the Blogosphere via Autotagging and Hierarchical Clustering. In Proceedings of the 15th international conference on World Wide Web, pages 625–632. ACM New York, NY, USA (2006)..

[Cacioppo99] Cacioppo, J. T. & Bernston, G. C. (1999). The affect system: Architecture and operating characteristics, Current Directions in Psychological Science, 8, 133-137.

[Cattuto08] Cattuto, C., Baldassarri, A., Servedio, V. D. P., & Loreto, V. (2008). Emergent Community Structure in Social Tagging Systems. Advances in Complex Systems 11 (4), 597-608.

[Couzin07] Couzin, I. (2007) Collective minds. Nature 445, 715

[Darwin65] Darwin, C. (1872/1965). The expressions of the emotions in man and animals. Chicago, IL: University of Chicago Press.

[Derks08] Derks, D., Fischer, A.H., Bos, A.E.R. (2008). The role of emotion in computer-mediated communication: A review, Computers in Human Behavior, 24 (3), pp. 766-785.

[Frijda86] Frijda, N. H. (1986). The emotions. Cambridge: Cambridge University Press.

[Forgas01] Forgas, J. P. (2001). Affect and the ‘‘Social Mind’’: Affective influences on strategic interpersonal behaviors. In J. P. Forgas, K. D. Williams, &L.Wheeler (Eds.), The Social Mind: Cognitive and motivational aspects of interpersonal behavior (pp.46-71). Cambridge: Cambridge University Press.

[Fortunato08] Fortunato, S. & Castellano, C. : Community structure in graphs. In Encyclopedia of Complexity and System Science. Springer, 2008.

[Giannakidou08] Eirini Giannakidou, Ioannis Kompatsiaris, Athena Vakali : SEMSOC: Semantics Mining on Multimedia Social Data Sources, Second IEEE International Conference on Semantic Computing (ICSC 2008), Santa Clara, CA, USA - August 4-7, 2008, IEEE Computer Society Press

[Katsaros09] D. Katsaros, G. Pallis, K. Stamos, A. Vakali, A. Sidiropoulos, Y. Manolopoulos: CDNs Content Outsourcing via Generalized Communities. IEEE Transactions on Knowledge and Data Engineering Journal (TKDE), Vol.21, N.1, pp. 137-151, 2009.

[Kelly01] Kelly, J. R., & Barsade, S. G. (2001). Moods and emotions in small groups and work teams. Organizational Behavior and Human Decision Processes, 86, 99-130.

[Liu, 2010] Liu, B.: In: Indurkhya, N., Damerau, F.J. (eds.) Handbook of Natural Language Processing, 2nd edn., Goshen, Connecticut, USA (2010)

[Newman06] Newman, M. E. J. Modularity and community structure in networks. Proceedings of the National Academy of Sciences of USA, 103, 8577–8582, 2006.

[Papadopoulos10] S. Papadopoulos, A. Vakali, and Y. Kompatsiaris: “Community Detection in Collaborative Tagging Systems”, Book Community-built Database: Research and Development, Springer, 2010.

[Papadopoulos11] S. Papadopoulos, Y. Kompatsiaris, A. Vakali, P. Spyridonos : “Community Detection in Social Media”. In Data Mining and Knowledge Discovery Journal, 2011.

[Radicchi04] Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., & Parisi, D. : Defining and identifying communities in networks. Proceedings of the National Academy of Science of the United States of America, 101, 2658-2663, 2004

[Raafat09] Raafat, R.M., Chater, N., Frith, C. (2009). Herding in humans. Trends in Cognitive Sciences, 13, 420-428.

[Tsagkalidou11] K. Tsagkalidou, V. Koutsonikola, A. Vakali, K. Kafetsios:"Emotional Aware Clustering on Micro-Blogging Sources", Proceedings of 2011 Affective Computing and Intelligent Interaction (ACII 2011).

[Nezlek08] John B. Nezlek, Konstantinos Kafetsios, C. Veronica Smith: Emotions in everyday social encounters: Correspondence between culture and self-construal. Journal of Cross Cultural Psychology, 39, 4, 366-372, 2008.

[Seger09] Seger, C.R., Smith, E.R., Mackie, D.M. (2009.Subtle activation of a social categorization triggers group-level emotions. Journal of Experimental Social Psychology, 45, 460-467.

[Shepitsen08] Shepitsen, A., Gemmell, J., Mobasher, B., & Burke, R. (2008). Personalized recommendation in social tagging systems using hierarchical clustering. In RecSys ’08: Proceedings of the 2008 ACM conference on Recommender systems, pages 259–266, New York, NY, USA.

[White05] White, S. & Smith, P. A Spectral Clustering Approach to Finding Communities in Graphs. In Proceedings of the SIAM Data Mining Conference (SDM), (pp. 76-84). Newport Beach, California, 2005

[Vakali08] Athena Vakali and Yiannis Kompatsiaris: Detecting and Understanding Web communities. In Proceedings of the WebSci’09: Society On-Line, 18-20 March 2009, Athens, Greece.