Analyzing Online Communities: A Narrative Approach

Andrew P. Feldstein
Department of Management and Marketing
School of Business
Virginia State University
P.O. Box 9209
Petersburg, VA 23806
Phone: (845) 367-2158
Email: Andrew.Feldstein@gmail.com

To be presented at MiT6 International Conference
Cambridge, Ma
April 24, 2009

Abstract

Marketing intelligence companies such as Nielsen's BuzzMetrics, J.C. Power's Umbria and Motivequest tell us that "there is gold buried in the mountains of data" accruing on blogs, online forums, and other forms of social media. The analytical services offered by these companies treat the various social media as one large database with which they decode "the language of the consumer" (http://motivequest.com). Semantic Network Analysis offers an alternative approach that "extracts and analyzes links among words to model an authors "mental map" as a network of links" (Carley et al. 2006). This paper suggests that this type of analysis can lead to a detailed and informative conceptual map of online conversations that will preserve the narrative context and offer a greater understanding of what motivates and holds these communities together.

The Internet has lead to a “radical decentralization of intelligence in our communications network...both the capacity to make meaning--to encode and decode humanly meaningful statements--and the capacity to communicate one’s meaning around the world, are held by and readily available to, at least many hundreds of millions of users around the globe” (Benkler 2006). This communication has taken many forms and each new technological platform leads to new and different forms of expression. Figure 2 showcases a fairly comprehensive sample of current techno-social platforms.

The rapid changes in technology, communications, and information exchange have been a boon to market intelligence companies who, when faced with this limitless amount of consumer-generated content have chosen to process all of it. Hence, Umbria, a part of J.D. Power and Associates “mines” consumer generated content from “hundreds of thousands of public conversations taking place in over 60+ million blogs, (as of October 2006) on message boards and other other online public forums” (umbrialistens.com 2009)

Umbria, a subsidiary of J.D. Power and Associates asserts “only once you identify who is saying what about your business when and where... can you determine how to effectively engage your audience” (umbrialistens.com). The Infegy website claims to have a social research application that harnesses “The world’s largest focus group” allowing marketers to “listen to millions in an instant” and thereby “know your market and consumers” (http://www.infegy.com). Other marketing intelligence companies warn of the danger in not listening to your customers. “Failing to listen to what customers have to say about your company can mean lost opportunities, lost sales, and even a damaged reputation (Zetainteractive.com 2009).

Converseon calls this process ‘conversation mining and asserts that this “is an essential first step in developing an effective communications strategy to join and influence the conversation (Converseon.com 2009). These marketing intelligence companies have all developed their own proprietary technologies to accomplish this goal. Buzzmetrics uses “innovative technology and rigorous methodology” in a 4-step process that involves “harvesting” various online sources, “cleaning” the data, determining its relevance, and then using “analytics” both technological and expert.

How this process works isn’t exactly clear. Umbria was the only company to answer my my email and their media representative explained that “proprietary reasons” prevented them from responding to the request for information. An explanation on the Brandwatch website comes closest to detailing an actual process. They start with custom search engine then use a “matching engine”; a form of artificial intelligence that uses specific algorithms to distinguish between Apple the brand and Apple the fruit.. Their “topic extractor then uses “Natural Language Processing” to extract common topics and a “Sentiment classifier” to determine how people feel about these topics. All of this creates a database that is massive in scope and available 24/7 saving hundreds of man hours (brandwatch.net).

But is it necessary, or even desirable to attempt to gather everything that consumers are saying about your brand? Is the information gathered from Facebook and Twitter going to yield as much information as the interactions among consumers who are part of an online brand community? Various social media platforms yield different types of information. If all of this information is aggregated will this, as the marketing intelligence companies promise, “ deliver a competitive advantage by uncovering and integrating data-driven insights culled from nearly 100

million blogs, social networks, groups, boards, and other CGM platforms" (Nielsen-online.com 2009). More to the point, will the information be truly meaningful?

While there is certainly something very tempting about being able to know what everyone is saying, a potential drawback of an analysis on this level and at this scale is that context will certainly be lost. In his article "New Media, Networking, and Phatic Culture" Miller decries the increasing usage of blogs, social networks, and microblogs and the concomitant 'flattening' of communication "towards the non-dialogic and non-informational" (Miller 2008). Viewed as isolated platforms one could certainly question the usefulness of 'phatic' communication limited to 140 characters per utterance that takes place on Twitter.

The same criticism could be leveled at Facebook or any of the other techno-social platforms that make it easy to contribute content that carries no more weight than would an ascenting nod or a disappointed groan. Conversations taking place in online communities, on the other hand, have the potential to contain considerably more content. Does it then make sense to amalgamate all of these messages and subject them to the same type of analysis? This is what J.D. Power, nielsen, converseon, and the others purport to do.

There is a distinct difference in the type of questions that can be asked or answered when your data is collected as dissociated bits. Typical marketing intelligence queries are concerned with brand reputations and perceptions that can be gleaned from the connection between a word representing a target brand and an adjacent adjective. If, for instance "Coke" is just a word or two away from "sucks" with alarming frequency this would certainly accomplish Converseon's goal of "providing early warning systems regarding brand or product issues" (converseon.com).

Umbria offers a competitive analysis that claims to provide information on attributes of competitive products that the market likes or dislikes. They also suggest that they can identify "what emotions are associated to competitive brands and products", "what consumer needs are unmet" and how product "is being compared to that of the competition" (Umbria.com). While these, and a whole host of other questions are undoubtedly important to brand managers and marketers, they are relatively simple questions. This is not to say that they are easy questions.

It takes a great deal of computational power to process the volume of data that Buzzmetrics and the other marketing intelligence use to fuel their proprietary databases and algorithms. However, if you closely examine the questions on their websites, you will discover that most of these questions are looking at very simple relationships; usually involving a dyadic relationship between products and attributes, products and target audiences, products and buzz, consumers and needs, emotions and brands, demographics and brands, and so on.

So, despite the large scale of the information being processed, marketing intelligence companies are working at a very basic level of analysis. In addition they seem to be ignoring some of the changes taking place among consumers because of the very technology that has generated all of this content. "Computer-supported social networks have brought a fundamental change in the nexus of human interaction ….. Operating as social networks has transformed community. Most community ties are now specialized, with different network members supplying emotional support, information, material aid, social identity, and a sense of belonging" (Wellman 2005). If we understand the specialized nature of ties it becomes easier to identify sources of information that are directly relevant to specific inquiries. Although this

process is more selective and directed than simply gathering all possible references to a product or concept, this can generate a potentially large and relevant sample target population.

The database approach does not provide information in its original context. To do this requires a very different approach to the collection and analysis of consumer generated content. If we can develop an approach which preserves the narrative context, more complex sets of relations will be available for analysis. The goal is to develop a method for processing relatively large amounts of consumer-generated content without losing the essential nature of the conversations nor the complex network of related concepts and ideas that comprise that conversation.

"Modern media is the new battlefield for the competition between database and narrative" (Manovich 2002). As a cultural form, database represents the world as a list of items and it refuses to order this list. In contrast, a narrative creates a cause-and-effect trajectory of seemingly unordered items (events). Therefore, database and narrative are natural enemies. Competing for the same territory of human culture, each claims an exclusive right to make meaning out of the world.

Theories and tools of analysis developed within the rubric of Social Network Analysis have given us tools to examine the structure of online communities. Mapping the organizational structure of brand communities will provide a starting point and a basis of comparison that will enable us to identify structural similarities and differences that exist between communities.

Key to this discussion will be a focus on the "imagined" aspects of online communities. The network perspective focuses on relationships and interactions instead of individual characteristics. "The relationship between a pair of units is a property of a pair and not

inherently a characteristic of the individual unit" (Wasserman and Faust 2006).

Social Network Analysis

"Relations are central to network analysis because they define the nature of the communication connections between people, groups, and organizations" (Monge and Contractor 2003). In describing relationships we look at characteristics that are quite different from the character attributes normally associated with individuals or organizations. Instead, it is the nature of the relational link or links that is important. The existence of a link or relationship will reveal some information but if we can determine the strength or directionality of that link we have a clearer picture of that relationship.

Network analysis "provides a collection of descriptive procedures to determine how the system behaves, and statistical methods to assess the appropriateness of the propositions" (Wasserman and Faust 2006). Network methods can be either descriptive or statistical. For the purposes of this study descriptive methods will be employed. These methods will provide formal measures that can be utilized to provide a substantive picture of more abstract, theoretical concepts. Escalas argues "that the meaning ascribed to products and brands can also be generated by narratives" (Escalas 2004). "What a brand means to a consumer is based in part on the narrative he or she has constructed that incorporate the brand" (Escalas 2004). ,

Online Communities

The first difference between these approaches involves the process of data collection. As discussed earlier, not every reference to a particular product is equally relevant nor equally meaningful. I will use Boston Red Sox to illustrate the point. A web search of online community

forums uncovers two online communities devoted to Red Sox fans; the Royal Rooters and the Sons of Sam Horn.

The Sons of Sam Horn has 1990 members who has produced over a million posts since the year 2000 (http://p100.ezboard.com/bsonsofsamhorn). The Royal Rooters has 6896 members who have produced more than 600,000 posts (start date not provided) (http://www.redsoxnation.net/index.php?act=idx). These are active communities and the members take their baseball very seriously. The conversations are in depth and chalk full of the language of true baseball fans.

Contrast this with the Red Sox fans group on Facebook which boasts over 600,000 members. The difference is that the Facebook fans don't talk about baseball and the extent of their contribution is simply clicking the yes/no button to affirm that they are, or are not, a fan. In this context you will see "Red Sox" and "fan" in close proximity numerous times but won't learn much else.

Active members in online brand communities tend to spend a lot of time interacting with other members. They engage in conversations, offer advice, complain, and otherwise share their thoughts by contributing copious posts to threaded discussions on community-based forums. According to a recent report from the Pew Internet & American Life Project , of the 71% of American adults that use the Internet "18% have posted comments to an online newsgroup or website" (Horrigan 2007)

The increasing participation by consumers in online communities of all types makes this community channel both interesting as well as crucial to the understanding of consumers within the context of the online community structure. By its very nature, an

online community is essentially different from a geographically bound, face-to-face social network. The concepts of 'space' and 'time', as well as the 'self' and the 'other' are the key elements that underscore this difference. While these differences have created limitations for online community interactions, they have also created opportunities.

"In the context of online communication, the social structure furthermore includes the technology's shaping of reality through its addition of an interrelated kind of structure. This structure is also about stable patterns or features; it constitutes the possibilities for interactivity and communication, the underlying ideologies of designs and organizations, connections and networks (Gotved 2006)." Although there is no 'physical' space in cyberspace, online community forums are created using software that has been specifically designed to give participants a sense of space through which they can navigate.

Crucial to the understanding of online community is the sense of commonality that members of a community feel toward one another. Muniz and O'Guinn (2001) describe the imagined community as a critical aspect of brand communities. "Consciousness of kind is the intrinsic connection that members feel toward one another, and the collective sense of difference from others not in the community" (Muniz and O'Guinn 2001). "Members' feelings of collective identity and common purpose (i.e., consciousness of kind) are dependent upon imagined commonalities to others sharing their consumer passion (also see Kozinets 2001; Schouten and McAlexander 1995). One key implication of the imagined community construct is that face-to-face interaction and physical proximity are not necessary for authentic experiences of community" (Thompson and Coskuner-Balli 2007).

Those joining online communities do so with some very specific expectations of what those characteristics are and how they will be manifest in community interactions. These expectations are based on a consumer's belief that those sharing an interest in a specific product or a particular brand also share other common characteristics. "Imagined commonalities are inculcated through public discourses and rituals that create a sense of common cause and identity (Thompson and Coskuner-Balli 2007).

Online communities are all about sharing. Members share opinions, advice, experiences, and stories. Through there posts a clear picture of the community begins to emerge. In the aforementioned Red Sox community it becomes very clear that these are not casual Red Sox fans. Their posts reveal an intimate knowledge of all aspects of baseball. Many of the posts on this discussion forum are cryptic in nature. Although I have been a Boston Red Sox fan all my life, there were numerous references that I had trouble deciphering. "MFY" was pretty obvious as a derogatory reference to the Yankees , but I wasn't quite sure what was meant by "You better not have "Russo-ed" this."

The fans who participate in this forum are hard core and there is little tolerance for lurkers (site newbies) who don't 'get it'. The "Yankees Elimination Magic Number" was created as a way for Red Sox fans to celebrate what would be the eventual elimination of the Yankees from the 2008 pennant race. One lurker named Kitchkinet mistakenly posted an image of the number '2' on the day after the Yankees elimination number was '3'. The Big Red Kahu, a Lifetime member responded "Hmmmmm… someone doesn't exactly "get" the concept here".Cuzitt followed with; "the next person who posts a "2" up in this thread while the Elimination Number is still "3" has a suspension headed their way".

Narrative

According to Mieke Bal, “a narrative is a text in which an agent relates a story” (Bal 1997). She goes on to make a clear distinction between the actual story and the text that is used to relate the story. Another distinction is made between the story and the fabula which “is based upon the difference between the sequence of events and the way in which these events are presented” (Bal 1997). If , as Bal does, we have are able to distinguish between these three levels of narrative in the stories that are related by a specific community of consumers then we can ask questions that are essentially different from the ones that are asked by marketing intelligence companies using text as a searchable database.

Bal uses the example of Tom Thumb. Why do we sympathize with Tom? Why do we cheer at the trick Tom pulls off whereby the giant consumes his own children rather than Tom? “This phenomenon demonstrates that something happens with the fabula that is not exclusively language-related” (Bal 1997). In the context of the Red Sox communities we might ask a similarly complex question such as; What does the Red Sox community’s portrayal of New York Yankee fans tell us about the structure and organization of the Red Sox community?

To answer the question we need to do more than just examine the attributes that describe how Red Sox fans feel about the New York Yankees and their fans. It is a multilevel question that goes on to ask whether the existence of the New York Yankees and their fans plays a formative role in the Red Sox community. To answer this question it is necessary to look at the complex interrelationship of concepts that community members use to tell the story of the community.

Semantic Network Analysis

A specific type of Network Text Analysis called Semantic Network Analysis "extracts and analyzes links among words to model an authors "mental map" as a network of links" (Carley, Diesner et al. 2006). The map analysis being used here differs significantly from more traditional content analysis methods. "Where content analysis typically focuses exclusively on concepts, map analysis focuses on concepts and the relationships between them and hence on the web of meaning contained within the text" (Carley and Kaufer 1993).

Therefore, we do not need to rely exclusively on the frequency of a concept to infer meaning. The network of concepts that is created through map analysis allows for statistical and graphical representation of the relationship between concepts. Because map analysis makes it possible to examine the data graphically and statistically, the researcher can stay close to the text and so augment qualitative techniques and capture the precision and inferential ability of quantitative techniques (Carley and Kaufer 1993).

Developments in Social Network Theory have produced a number of theoretical perspectives and innovative tools with which to examine online brand community interactions. "Comparison and exploration of abstract data are much easier when mapped into a visual structure: conceptual similarity may be visualized as shape or color similarity, or as spatial proximity" (Abrams and Hall 2006). Applying Network Text Analysis to an online community's threaded discussions can render a detailed and informative conceptual map of the relations between the concepts being discussed by community members. Social Network measures such

as centrality will be used to gauge the power, strength, and influence of concepts and the relations between and among them.

In a graphical representation of online community threaded discussions, community members are identified as nodes. "Edges" are the lines that convey information about the relations between nodes. A well-rendered graphical representation, or visualization of community member attributes and the relationship between these members can present a picture of the network that can lead to intuitive revelations about the social structure and nature of the community.

The creation of a visual map of online threaded discussions, in combination with relevant network analysis measures, provides a summary overview of what is being discussed by online community members. The representation of concepts within a forum community discussion are a direct reflection of how members of that community perceive those concepts. In attempting to understand online communities it is necessary to be able to 'see' concepts as they see them. Subtle, and not so subtle, differences in the way a particular concept fits into the structure of a threaded discussion can indicate the relevance and importance of that concept to a given online community.

Texts can be coded and analyzed as networks of concepts often referred to as maps or semantic networks. In such networks, for many texts there are elements of social structure - the connections among people, organizations, and events. (Diesner and Carley 2005) "Within a cognitive map, the meaning of a concept is the aggregate set of relations it has to all other concepts that make up a conceptual network. Mental models are dynamic structures that are constructed and expanded as individuals make inferences and gather information. They contain

both specific information about particular items, and general (or social) knowledge. A transcript of an individual's speech is a reflection of the individual's mental model at a particular point in time. Accordingly, such texts may be thought of as a sampling of information from an individual's memory" (Popping 2000).

Articulating the Narrative

To understand what consumers are saying (story), the structure of their conversations (text), and the logic of the events being related in that story (fabula) it is necessary to develop a method which can accommodate structure as well as meaning. I believe that a method built around Network Text Analysis and social network measures will bring us closer to that goal because it is the narrative aspect of consumer-generated content that will help to identify complex patterns and interrelationships.

Bibliography

Abrams, J. and P. Hall, Eds. (2006). Else/Where: Mapping New Cartographies of Networks and Territories. Minneapolis, Mn, University of Minnesota Design Institute.

Bal, M. (1997). Narratology: Introduction to the theory of narrative. Toronto, University of Toronto Press.

Benkler, Y. (2006). The Wealth of Networks: How Social Production Transforms Markets and Freedom. New Haven, Yale University Press.

brandwatch.net. "Brandwatch Technology: The key strength of brandwatch is the technology it uses." Retrieved January 30, 2009, from http://www.brandwatch.net/technology.

Carley, K. M., J. Diesner, et al. (2006). AutoMap User's Guide, Carnegie Mellon University (ISRI) Institute for Software Research International(CASOS) Center for Computational Analysis of Social and Organizational Systems**:** 110.

Carley, K. M., J. Diesner, et al. (2004). An Integrated Approach to the Collection and Analysis of Network Data, (CASOS) Center for Computational Analysis of Social and Organizational Systems Carnegie Mellon University**:** 9.

Carley, K. M. and D. S. Kaufer (1993). "Semantic Connectivity: An Approach for Analyzing Symbols in Semantic Networks." Communication Theory **3**: 30.

Converseon.com. (2009). "Conversation Mining: Are Your Ears Buzzing ?" Retrieved January 30, 2009, from http://www.converseon.com/conversation-mining.html.

Diesner, J. and K. Carley (2005). Revealing Social Structure from Texts: Meta-Matrix Text Analysis as a novel method for Network Text Analysis Causal Mapping for Information Systems and Technology Research: Approaches, Advances, and Illustrations. V. K. Narayanan and D. J. Armstrong. Harrisburg, Pa, Idea Group Publishing**:** 81-108.

Escalas, J. E. (2004). "Narrative Processing: Building Consumer Connections to Brands." Journal of Consumer Psychology **14**(1/2): 168-180.

Gotved, S. (2006). "Time and Space in Cyber Reality." New Media and Society **8**(2): 467-486.

Horrigan, J. B. (2007). A Typology of Information and Communication Technology Users. Pew Internet & American Life Project. Washington D.C., Pew Internet & American Life project.

Kozinets, R. (2002). "The Field Behind the Screen: Using Netnography for Marketing Research in Online Communities." Journal of Marketing Research **39**: 61-72.

Manovich, L. (2002). The Language of New Media. Cambridge, MIT Press.

Miller, V. (2008). "New Media, Networking. and Phatic Culture." Convergence **14**(4): 387-400.

Monge, P. R. and N. S. Contractor (2003). Theories of Communication Networks. New York, Oxford University Press.

Muniz, A. M. J. and T. C. O'Guinn (2001). "- Brand community." Journal of Consumer Research **27**(4).

Nielsen-online.com. (2009). "nielsen online." Retrieved January 9, 2009, from http://www.nielsen-online.com/products.jsp?section=pro_buzz&nav=1.

Popping , R. (2000). Computer-assisted Text Analysis. London, Sage Publications.

Thompson, C. J. and G. Coskuner-Balli (2007). "Countervailing Market Responses to Corporate Co-optation and the Ideological Recruitment of Consumption Communities." Journal of Consumer Research **34**(2): 135.

umbrialistens.com. (2009). "Web Intelligence powered by Umbria." Retrieved January 3, 2009, from http://www.umbrialistens.com/services/index.php

.

Wasserman, S. and K. Faust (2006). Social Network Analysis: Methods and Applications. New York, Cambridge University Press.

Zetainteractive.com. (2009). "Zeta Interactive: Your Customers are Talking Are You Listening." 2009, from http://www.zetainteractive.com/what_we_do/social_media_mining.aspx.