

promoting access to White Rose research papers



Universities of Leeds, Sheffield and York
<http://eprints.whiterose.ac.uk/>

White Rose Research Online URL for this paper:

<http://eprints.whiterose.ac.uk/74462/>

© 2010 IEEE. Reprinted, with permission, from:

McDermott, SE (2010) *White's Three Disciplines and Relative Valuation Order: Countering the Social Ignorance of Automated Data Collection and Analysis*. In: Memon, N and Alhajj, R, (eds.) Proceedings of the 2010 International Conference on Advances in Social Network Analysis & Mining. 2010 International Conference on Advances in Social Networks Analysis and Mining, 09-11 August 2010, Odense, Denmark. IEEE Computer Society , 72 - 79 .

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the products or services on this website. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

White's Three Disciplines and Relative Valuation Order

Countering the social ignorance of automated data collection and analysis

Steven McDermott (Author)
Institute of Communications Studies
University of Leeds
Leeds, United Kingdom
S.E.McDermott@leeds.ac.uk

Abstract—This paper asks which of White's (2009) three disciplines and relative valuation orders does the Singapore blogosphere adhere to. Analysing not just the hyperlink connections but the textual discourse; and in doing so attempts to highlight certain limitations of using automated data mining and analysis software. Using the Singapore blogosphere, described by Lin, Sundaram, Chi, Tatemura, and Tseng, (2006) and Hurst (2006), as an isolated and distinct network with no theme or focus, I have targeted blogs using social network analysis uncovering the key players, with higher levels of 'betweenness centrality' (de Nooy & Mrvar et al., 2005) and the themes and discipline of the Singapore blogosphere. This case study will help highlight the analytic framework, benefits and limitations of using social network analysis and an ethnographical approach to networks. This paper also highlights the use of various software technology; blogs, *IssueCrawler*, *HTTrack*, *NetDraw*, and *Leximancer* while using an ethnographic approach to counter the social ignorance of automated electronic software.

Keywords – social network analysis; semantic network analysis; social ignorance; data mining; disciplines

I. INTRODUCTION

This paper demonstrates the use of *IssueCrawler* (Rogers, 2006) a Web 'mapping' device to conduct hyper-link network analysis and visualise the blogs orientated towards discussion of matters concerning Singapore. I target blogs that are engaged in discourse regarding Singapore or have the author geographically living in Singapore. This shadow of the entire network produces the URL seeds for the extraction of the textual data using a site scraper called *HTTrack* (Roche et al., 2007). The data is then initially analysed using *Leximancer* (Smith & Humphreys, 2006). *Leximancer* analyses the textual data and extracts two types of data, semantic and relational; it is an automated content analysis process. This then enables me to visualise the discourse beyond my immediate and limited scope of awareness as an ethnographer in the Singapore blogosphere.

This paper sits within the wider debate regarding the emancipatory nature of the Internet in authoritarian regimes (Habermas, 2006), with Singapore as a case study.

It is rarely that we find a community that is absolutely isolated, having no outside contact. At the present moment of history, the network of social relations spreads over the whole world, without any absolute solution of continuity anywhere. This gives rise to a difficulty which I do not think that sociologists have really faced, the difficulty of defining what is meant by the term 'a society'. They do commonly talk of societies as if they were distinguishable, discrete entities [...] is the British Empire a society, or a collection of societies? Is a Chinese village a society, or merely a fragment of the Republic of China? (Radcliffe-Brown, 1940:224)

What are the methods of control and what is the nature of the Singapore identity online? These two questions are analysed by interpreting the first question as a question relating to 'power' within the blogosphere. The nature of the Singapore identity is the identity that appears in an analysis of the discourse. Identity emerges out of attempts of control in a context of constant and turbulent flux (White & Godart, 2007).

II. SOCIAL NETWORK THEORY

Social order develops from the interplay of social structure and culture amid uncertainty. Social structure and culture are part of the complexities of discursive practices of those who populate the institutions and organizations. As White (2008) states, *identities* are the results of attempts at control whilst within the chaotic and entangled generative mechanisms of culture and social structure. It is these attempts at control that become the social realities for other identities, this does not entail domination, coercion or force. One identity attempting

to gain a root does not result in the uprooting of other identities. Others in an unproblematic manner assign continuity to that social reality even though that social reality is adding to and maintaining the chaos that surrounds them and works in order to constitute them. We wrongly assign normality to the context. This 'assigned normality' contrasts with the haphazard and chaotic nature of what we directly experience as human beings. There is an ongoing haphazard bringing together of various perceived or reinterpreted components of the normal in order to assert an identity and thereby a sense of control where none exist. These identities are the sources of and receivers of communication to which identities assign meaning (White, 2008).

Meaning roots identities to stable positions from which information emerges. The accessing and assessing that passes between the identities are the entangled chaotic mix of culture and social structure from which meanings emerge. *Switchings* of meanings happen in physical and social environments that are uncertain that result in identities attempts at acting across and amongst divergent networks and domains of topics. It is this switching by which identities generate meanings and discursive formations or styles. Mische and White (1998) assert that *netdoms* are first order elements from which networks and domains derive. Network relations and discursive formations are inter-related and co-constitutive. Networks are the formations of stories and stories are the formation of networks. "Social networks and discursive formations are second – order processes which need to be accounted for from the dynamics of identity and control among netdoms" (White, 2008). New or fresh meaning emerges for humans during the process of switching between netdoms.

Netdoms are not things or physical phenomenon but are experiential processes that are so overwhelming that the individual within it is unable to bring it into focus (White, 2008:7). Networks are identities and a common set of stories that are able to explain anomalies away and provide the space for social action to happen. Disciplines emerge out of network forms as the struggle for control comes to rest on a *valuation order*.

According to White (2008), there are three genres of disciplines; interface, council and arena. Each genre has a distinct process and valuation order. Interface discipline includes the flow of production that ensues when something physical is being made, but also the conduct of lectures in a university or the production of toys. It requires a level of commitment from the actor, whether that is people or other groups and the valuation order is that of quality. This can be quality in a publication, fame, number of citations, number of graduate students or finished product. Council discipline is applicable to university councils or committees that have dominance over a particular resource. The process that takes place is that of mediation in order to regulate the flow of people. The valuation order is that of prestige in particular

linked to the influence of corporate action requiring 'soundness' of character. The arena discipline is the formation of order that surrounds groups such as country clubs, the American university fraternity house, or the Rotary Club, with a distinct focus on selection via boundaries of inclusion and exclusion in order to uphold a sense of 'purity' within the perceived group, passed on via marriage or eligibility for marriage. Each of these disciplines is a status system where evaluative judgments create a network pattern. The judgments that these disciplines make create task flows that generate network patterns.

These local status systems co-ordinate differently but are primarily geared towards getting things done. The networks that form around an interface discipline pumps flows from the outside upstream to downstream outside; the council discipline creates flows that reach out in attempts to control resources and people, while the arena discipline creates boundaries to avoid introducing perceived impurities into the group. The "participants commit to producing flows in interface, whereas in council they mediate among proposals, and in arena they select from candidates" (White, 2008:65). White's concepts are not used as hard and fast descriptions but heuristic guides to observation. Hard and fast descriptions would run contrary to the position that neat and precise discriminations amongst social and network structures is at odds with more stochastic measures of messy social life.

The embedding of a discipline can adhere to one or a combination of formats, involution, dependence or differentiation. Involution of a discipline among other disciplines occurs when embedding takes place amongst the ricolochets from network processes in and around other disciplines. Involution is when the chains of ties flow back in on their origin, which ensures the continuation of the valuation order – *specialization*. Embedding via dependence occurs when trying to accomplish a joint task in the operational environment, the discipline becomes locked together with external disciplines, this becomes evident when a particular discourse in stories and physical activities *interdigitates*. Embedding via differentiation happens when appropriate valuation members focus their attention towards those with a perceived higher prestige and begin to attract attention from outside the discipline, the differentiation determines a level of *visibility*.

Valuations are observable by inference and induced by looking at the story-sets active in the localized discourse. These story-sets contain within them the valuation order that has the potential to become hegemonic. At this stage, the discipline becomes an actor. Valuation orders also embed but do so internally within the discipline. Each member of the group ascribes to the valuation order. Disciplines require orderliness of at least perceptions by members of the discipline. The control struggles that formed the discipline surround the identities, settle into networks and settle on a

particular style, institution and regime (Padgett & McLean, 2006).

White's conceptualization of disciplines raises the question 'can the Singapore blogosphere become a control regime in conflict with the already embedded control regime of the Singapore state if it is enmeshed in the rhetoric, narrative and style of the Singapore state?' Or rather - Is the Singapore blogosphere challenging the hegemony of the Singapore state?

III. SOCIAL NETWORK ANALYSIS

I chose the Singapore blogosphere as a case study as it is an isolated and distinct network with no theme or focus according to Lin, Sundaram, Chi, Tatemura, & Tseng, (2006) and Hurst (2006). In adherence with Nadel (cited Cavanagh, 2007), the important factor is to not to arbitrarily demarcate a unit of analysis to study. By reducing the social to the network, it allows for the unit of analysis studied to materialise throughout the process of conducting the research. The network is discovered through empirical evidence rather imposed at the beginning.

Social Network Analysis is a formal, mathematical technique of analysing relational data. It is concerned with the contacts, ties and connections, group attachments and meetings (Scott, 2000:3). "The relations are not the properties of the agents themselves, but of systems of parts; these relations connect pairs of agents into larger relational systems." (Scott, 2000:3). The appropriate method for the analysis of relational data is network analysis of qualitative measures of network structures. Emphasis is on the 'structure' of social action. "Structures are built from relations" (Scott, 2007:4). Social meaning constructed by the group members of the network is founded on the perceptions and experiences of the context in which they are operating. Paths of connections run between the groups and these paths divide the groups into distinct regions. A lack of paths separates regions (Scott, 2007:11). Paths run within the regions but not between the regions. The regions are constraints or boundaries. These boundaries are the 'forces' that determine group behaviour.

Deleuze and Guattari (2004) present a model of knowledge and perception known as *rhizome*. The rhizomatic model of knowledge results in a network model that appears to be chaotic. Rhizomatic networks mean that a path to every other point in the network connects any point in the network. The logic of the connection in the rhizomatic network is movement. A connection is the sprouting off in a new line. The lines may appear to be random as they do in hyperlink analysis but they do have a purpose. Hyperlinked culture has as its main aim 'intertextual evolution' (Dreyfus, 2001) whereby all possible associations and linkage is enabled regardless of how tenuous they may appear. Resulting in a disordered knowledge and enabling a new form of knowledge to emerge. The main point with Deleuze and Guattari's rhizomatic network is that there is no hierarchy; no node takes

precedence over another. The order is in constant flux with total inclusiveness. The flow of information however pre-dates the existence of the nodes. The nodes are interruptions in the flow. The nodes merely channel the flow of information. The Internet is the node in the flow of information and knowledge exchange (Cavanagh, 2007).

Knox et al. (2006) argue that American Social Network methods map roles comprehensibly and this results in the incorrect assumption that they have delineated the 'real' social structures. This ends up reinforcing a view of relations that are very far removed from the everyday experiences of people. Knox et al. argue that Social Network Analysis' focus on structuralism has in recent years shifted to attempts at developing a cultural approach.

IV. NETWORK ETHNOGRAPHY

Here I present the suitability of applying the ethnographic approach to the study of the Singapore blogosphere. My major concern with research conducted on the Singapore blogosphere by Lin, Sundaram, Chi, Tatemura, and Tseng, (2006) and Hurst (2006) is that it led to technological and/or organizational determinism.

Howard (2002) has put forward what is viewed as an amalgamation of social network analysis and ethnography for the study of 'new media' including 'epistemic communities'. Howard argues that social network analysis is better at defining a core group of members of a group and expanding on that number than traditional ethnographic approaches. Network ethnography is the process of using ethnographic field methods on sites or nodes selected by social network analysis including an online network such as the Singapore blogosphere. According to Howard (2002), "network ethnography allows the qualitative researchers to think strategically about the selection of cases by empowering them to define the universe of cases themselves". Howard (2002) also argues that such an approach will undermine inherent problems of the qualitative approach such as sample bias and maintain a balance between technological and organizational determinism on the one hand and the social construction of culture on the other hand.

V. EXTRACTING THE SOCIAL NETWORK USING HYPERLINK NETWORK ANALYSIS

Social network analysis traces the flow of information that passes through a network of relations. As actors make use of computer networks the computing networks are "clear indicators of communication structures within society" (Garrido & Halavais, 2003). Garrido and Halavais posit that, "A map of the communication network is roughly isomorphic to the structure of the relationships among the users (2003)."

In hyperlink analysis, the unit of analysis can be a Website, a hyperlink, a blog. The unit of analysis are blogs and the

hyperlinks between them. Hyperlinks allow actors or groups of actors to form social and communicative relations. Jackson (1997) suggests that the methods of social network analysis maybe useful in understanding the relations of computer mediated social processes. Hyperlinks are not just technological tools but social channels. Web sites and blogs are actors with hyperlinks among blogs representing a relational link (Park, 2003). A hyperlinked system exchanges information and cooperates around a given shared background, interest or project.

To begin I initiated a hyperlink analysis using *IssueCrawler* (Rogers, 2006) a Web ‘mapping’ device in order to conduct hyperlink network analysis and visualise the blogs orientated towards discussion of matters concerning Singapore. It is publically accessible software offered by the Amsterdam based *Govcom Foundation*. Bruns (2007), Siapera, (2006) and McNally (2005) have all used *IssueCrawler* to uncover web-situated networks. *IssueCrawler* is a server side Web network location software (Rogers, 2006). Here I outline my specific use of *IssueCrawler*. For a more detailed account, see Bruns (2007). *IssueCrawler* requires the user to input URL addresses as seeds and it runs a crawl of the URLs capturing page and site out links. It then performs co-link analysis, which means that it checks to see if the site that links to another has that link reciprocated.

An important factor in beginning a crawl-based analysis of a blogosphere is the starting points or URLs. According to Park (2003), there are two methods for gathering data on hyperlink networks. The first involves the researcher observing the sites and the second uses computer-assisted measurements. The use of human coders in the observational method does allow error to enter the process; it could also involve high labour costs. I used the second method of computer-assisted measurements for this research. I started by gathering a large set of political blog addresses, or URLs, by downloading a list of political blogs¹ compiled by a group of Singapore bloggers. I then expanded this of URL addresses or seeds using a snowball approach that combined an initial list of sites grouped along themes such as ‘socio-political’, ‘gay and lesbian’, ‘social’, ‘Christian’, and ‘Malay speaking’. Using *IssueCrawler* I harvested the URLs and used them as start nodes for the hyperlink analysis in July of 2006. The parameters were set as follows; number of iterations – 2; crawl depth – 2; co-link analysis by page and

privileged starting points on. (This keeps the URLs in the results after the first iteration.) Over a period of twelve months, this resulted in 1,220 blogs and websites.

I relied on two types of data output generated by *IssueCrawler*. I used the UCInet data file. The UCInet data file is a *NetMiner* compatible output file that allowed me to conduct my own social network analysis. The second output file was the *IssueCrawler* xml file that contained the list of URLs visited by the web crawler in order to conduct the *HTTrack* site extraction.

I use *Netdraw* (Borgatti, 2002) for the social network analysis and visualisation of the Singapore blogosphere. I used measures of closeness centrality, closeness centralisation, betweenness centrality and betweenness centralisation to assess which blogs are more ‘important’ or rather in ‘control’ of the flow of information. The closeness centrality score of a particular blog indicates the number of other blogs divided by the sum of all distances (the shortest path between two blogs) between the blog and all other blogs (de Nooy & Mrvar et al., 2005). Closeness centralisation, which refers to the entire network or blogosphere, is the variation in the closeness centrality of blogs divided by the maximum variation of closeness centrality scores possible in a network of equal size (de Nooy & Mrvar et al., 2005). Fig. 1 shows the results of a 2-faction analysis of the data sets for January 2009 and February 2009. A blogs position indicates whether it has access to information and better opportunities to spread information. As the network of blogs becomes less centralised the possibility of the information flow being distorted increases. A method of assessing the ‘importance’ of a blog in the network is to measure how large a role it plays as an intermediary (de Nooy et al., 2005). How many flows of information are disrupted when a blog becomes inaccessible or inactive? How many detours are required to access other blogs, which blogs control the flow of information because of their position in the network, if a blog is taken out of the network? To paraphrase de Nooy et al. (2005) “The

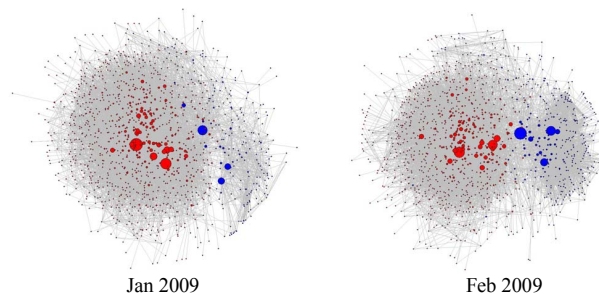


Fig.1 Singapore blogosphere from Jan and Feb 2009 divided into 2 factions. The size of the node refers to in betweenness centrality. The red circles to the left are blogs written in English and the blue circles on the right are blogs written in Malay.

betweenness centrality of a [blog] is the proportion of all hyperlinks between pairs of other [blogs] that include this [blog].

Betweenness centralisation, which again refers to the entire network [or blogosphere], is the variation in the betweenness centrality of [blogs] divided by the maximum variation in the betweenness centrality scores possible in a network of the same size.”

¹List available <http://spreadsheets.google.com/pub?key=pQcRq80yuyWXeqpaklg2ibA>

VI. CORPUS DATA COLLECTION AND ANALYSIS

After extracting the list of crawled URLs from the *IssueCrawler* results, I used them as input addresses to extract the front pages of each website and blog using *HTTrack*. The Internet mirroring created files of approximately 12 megabytes in January and 12 megabytes in February 2009. It is the equivalent of 4,000 pages of textual data on both occasions.

Analysing such a large corpus introduces certain questions of scale and time. Manually attempting to analyse so much data would also increase the likelihood of human error and the potential bias of the researcher might skew the results. While the solution offered here limits the ‘in-depth’ analysis available to an ‘off-line’ ethnographer I feel that the results generated by the software programme still require an assessor who is able to construct a narrative that fits with the software generated results and what has actually happened. Common approaches to analysing such large data sets would be to conduct a content analysis. The software employed here begins by conducting a content analysis, assessing the key themes within the corpus it then assesses the concepts that co-occur with that theme. As computer software is socially ignorant, the ethnographic component to this approach enables a social reading of the results to ensure that themes highlighted by the software make sense in the given context. The resulting co-occurrence information creates a concept map that is then read by the researcher.

Smith and Humphreys (2006: 262) state that by using *Leximancer* the analyst is aware of the global context and significance of the concepts and helps ensure that the analyst does not become fixated with some concepts to the detriment of others. *Leximancer* (Smith & Humphreys, 2006) is a;

Method for transforming lexical co-occurrence information from natural language into semantic patterns in an unsupervised manner. It employs two stages of co-occurrence information extraction—semantic and relational—using a different algorithm for each stage. The algorithms used are statistical, but they employ nonlinear dynamics and machine learning.

Leximancer uses a combination of techniques such as Bayesian statistics that records the occurrence of a word and connects it to the occurrence of a series of other words. It then quantifies those outputs by coding the segments of text, from one sentence to groups of sentences or entire paragraphs depending on what the user has requested. Each word or concept is associated with a subset of related terms. The next step involves the machine learning from the concepts already uncovered and linked to other concepts creating a ‘concept space’. It then iteratively creates a thesaurus around a group of seed concepts. This information is visualised using network analysis. Emergent themes are then visible to the user, and are expandable using the map visualisation that links directly to

the areas of the corpus in which the concept occurs. The emergent themes enable a quick reading of the large data set. For my purposes, it enables me to see what the dominant themes are, rather than imposing my own on the data, and helps me to navigate the large data set. The ethnographic component of my research then enables me interpret the themes and concepts according to my accumulated knowledge gathered as part of being a participant observer in the network for 5 years.

VII. ANALYSING ETHNOGRAPHICALLY ACQUIRED KNOWLEDGE

Here I consider issues around the notion of *A)* ‘being in context’ *B)* time or presence and absence, *C)* ethical considerations, and *D)* network ethnography and online versus offline approaches to ethnography, by blogging or rather producing an online web presence. The observer aspect of the method was the textual data collection process. As well as the limitations and benefits of various forms of methods used in online research there are also theoretical and methodological issues to be considered which are touched upon when looking at the problems of technological and organizational determinism. Ethical considerations are central to any ethnography and the current geographic and online nature of the network highlights certain concerns.

A. Being in Context

A problem encountered when in the field is that of defining the boundaries of the field or network. This problem is not unique to the study of online *virtual* networks. As Atkinson, (1992) states the field is produced (not discovered) through the social transactions engaged by the ethnographer. The boundaries of the field are not given. They are the outcome of what the ethnographer may encompass in his or her gaze. The nature of a network in a physical setting implies that it is socially constructed and so the reference to online networks as *virtual* networks is posited on the notion and belief that there are *real* physically posited networks to make comparisons with. According to Renninger and Shumar (2002), there are three features in a network needed to be apparent in order for it to be a virtual network. Firstly a core group of users who continually return to a particular site of online interaction need to be identified, secondly temporal and spatial possibilities need to be considered and thirdly the linking of conversations between websites, archiving discussions and thereby allowing for the possibility of future discussions around the same resources.

Ethnographers in the ‘traditional’ sense are able to get out there and spend time learning about people in their native or natural context. The online ethnographer will encounter the same obstacles or learning paths such as gaining access, maintaining relationships.

B. Textual Presence and Absence

The participant aspect was carried out by blogging or rather producing an online web presence within the network highlighted by Lin, Sundaram, Chi, Tatemura, and Tseng (2006) and Hurst (2006). Such an undertaking includes issues that are not particular to online research. Rutter and Smith (2005) have referred to the automation of most of the data collection process for a technically aware social researcher. However, they argue that a sustained presence of the ethnographer in the culture under study is essential. As texts are the key elements in ethnography and within textual anthropology, in this instance textual data forms the raw data. The researcher as a participant in this instance seems to be in the generated text produced by the ethnographer, whether that is blog posts or comments on the sites within the network.

The representation of the online network or blogosphere is at the textual level, and can only exist at the textual level. It will however be asserted that interpretations of texts by researchers and analysts maybe helpful. Ethnographic work and analysis of the discourse used allows the blog producers' aims and how readers use the texts to be got at.

Online research requires the researcher to become a member of a core group of users who continually return to a particular site of online interaction. Regular or daily interaction is required in order to fit into the daily routine of other participants. Interacting daily within the group enables learning and understanding as an outsider. Enabling the researcher to add to questions and follow alternative directions of research. Being able to blend in and acquire a feel for the norms of behaviour of the network is also crucial. In order to become a member of the online network I created a blog that other bloggers could create hyperlinks to. Resulting in high betweenness centrality scores for the blog.

Data archiving of texts, was automated and currently stands at 24 megabytes for the two-month period. The habits, customs and myths although made manifest in textual data will require the understanding of the technological and organizational setting as well as the culture being created and re-created by the people who populate the Singapore blogosphere. This is an attempt to overcome the social ignorance of automated data collection and analysis software.

Several issues are however to be considered when conducting online ethnographic research. As well as the suitability of various forms of methods used in online research there are also methodological issues to be considered.

C. Ethics of Online Research

Although the blogosphere is in the public domain and the researcher is exempt from gaining informed consent, Reid (1996) argues that researchers should ensure to minimize the harm caused during the process of dissemination from 'disinhibited exposure' that many feel on the Internet. With the

'authoritarian' nature of the Singaporean state well documented these ethical concerns and guidelines will have to be carefully adhered to before the dissemination of the research findings.

According to Hine (2005), there is considerable anxiety about how far traditional research methods are appropriate for studying technologically mediated interaction. As with offline ethnography, the issues of misrepresentation, consequentiality and anonymity play an important and key role in the present research design. It is also important to understand the wider offline environment in which the ethnographic research is conducted when considering ethical issues.

The Internet facilitates existing social trends while at the same time opening efficient modes of social control (Lyon, 1994; Lyon & Zureik, 1996). An ethnographic approach will make the views and attitudes of the day-to-day users of the technology explicit and uncover how they interpret the engendering and endangering repercussions of engaging with the emergent properties of the blogging network. As levels of social control are high in Singapore it is paramount that I protect the anonymity of those involved in the various daily interactions throughout the process of conducting the research.

I also consider the repercussions of being a 'lurker' when merely engaged in observation and the issues of copyright law when archiving the particular documents posted by various members of the blogosphere. There is also the issue of 'researcher effect' to be considered when actively engaged with the Singapore blogosphere. The disclosure of my own identity as a researcher has occurred, which may diminish validity claims.

The researcher should not cause harm and public dissemination of the actual names or IP addresses of the participants would infringe such ethical guidelines.

I have been an active participant in the field as any traditional ethnographer would be. This process of negotiating the field from an ethical position will be an aspect of the texts generated by the researcher when making decisions and postings within the online network.

D. Research Offline and Online

Trying to generate a holistic approach to ethnography according to Hine (2000) is difficult as every account is always selective and partial. However, Wilson (2006) has taken the approach that the way forward maybe discerned through an "integration of ethnographic methods, both traditional (offline face-to-face) and virtual", as they "can be helpful in developing rich and comprehensive understandings of relationships between online and offline [...]" (Wilson 2006:309). However, this in no way undermines the importance of ethnographic approaches to Internet research as stressed by Markham (1998), Miller and Slater (2000), Mann

and Stewart (2002, 2003), and Kendall (1999), as well as Hine (2000).

VIII. THEMES AND DISCIPLINES

The use of *Leximancer* highlights the themes and concepts that aid in the scope of awareness that I have as a participant observer within the network. It increases my scope of awareness as an observer, well beyond my own ego-network of daily interaction. However the use of *Leximancer* is not a quick fix for trying to assess the discursive content of large scale networks, the skill of ‘reading’ the maps requires an awareness of the social context in which the data was gathered. Stockwell and Colomb et al., (2009: 436) assert that “[a]s users develop in expertise; it can be used to explore subtle aspects of more familiar domains. As with any automated system, skill in its use lies in applying it to problems for which it is suited”, but the expertise required is not merely one of the user working out how to use the software but one of overcoming the social ignorance of the software itself.

Fig. 2 is a screen grab of results from the data collected in January 2009. It enabled me to assess the dominant themes and related concepts that were in use in the Singapore blogosphere for that time. The dominant themes were *dengan* (*with*), *comment*, *people*, *day*, *Singapore*, *time*, *public*, *txt*, *unregistered*, *New Year*, *political* and *pm*. The first concern raised in the use of *Leximancer* was the format of the data as it is put into the software. The theme titled *unregistered* is not in the corpus that had been collected from the Singapore blogosphere but had been a component of how the data had been stored after collection (converted to a pdf format). *Leximancer* however allows the user to remove such themes by creating a thesaurus. The most dominant theme *dengan* (Malay for ‘with’) that registered 986 hits also highlighted my limitations, at the time of conducting the research, as I was unaware that it could perform word co-coherence in multiple languages at the same time. The appearance of *comment* as a theme merely highlights the dominance of the use of comments in blogs with the majority of blog posts possessing

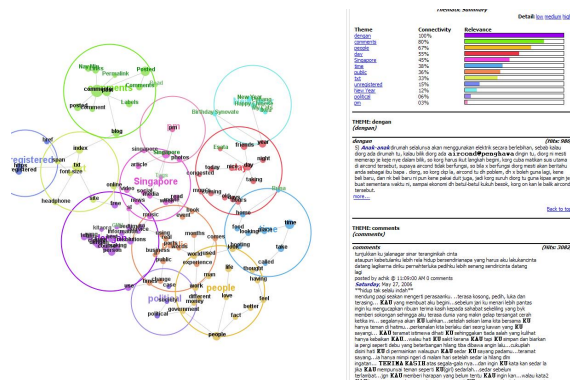


Figure 2 Screen grab of results of Leximancer word co-coherence of the corpus of data gathered in January 2009 of the Singapore Blogosphere.

that functionality. The appearance of themes such as *Singapore*, *public*, and *political* are because of high occurrence of concepts appearing that were generated by the thesaurus. *Leximancer* provides a list of the appearance of those words in the corpus, which as a participant observer of the network is useful. It highlights story sets that I was aware of as well as numerous story sets that I was not. A more detailed look at the theme *Singapore* brought to my attention bloggers that I was unaware and concepts that were associated with Singapore, such as ‘*Nation Cheated*’. What *Leximancer* enables is a quick jump from themes to text segments. It enables rapid jumps from concept co-occurrences in map format, at a global scale, to how the terms are used in the context of specific segments of the text (Stockwell & Colomb et al., 2009). The speed and effectiveness that *Leximancer* provides is commendable. It is however software that requires a detailed understanding of the social, cultural and political nature of the environment under scrutiny.

IX. CONCLUDING REMARKS

The ethnographic approach is integral to unpacking the social and cultural forces that are at work *within* and *without* the network in order to counter the social ignorance of automated electronic software.

TABLE I. WHITE’S (2008) THREE DISCIPLINES AND RELATIVE VALUATION ORDER

| Discipline | Interface | Council | Arena |
|-----------------|--------------------------------------|-----------------------------------|--------------------------------|
| Valuation Order | Quality – e.g. University Lecture | Control – e.g. University Council | Purity – e.g. Rotary club |
| Nature of Flow | Outside upstream, downstream outside | Reaches out to control | Creates Boundaries |
| Embedding | involution via specialisation | dependence via interdigiting | Differentiation via visibility |

White’s (2008) concepts are not hard and fast descriptions but heuristic guides to observation that enable stochastic measures of messy social life. However, this case study covering two months of corpus highlights a Singapore blogosphere embedded via dependence as it has become locked together with the external disciplines of the Singapore government and government affiliated mainstream press. The particular discourse in stories and the topics discussed has become *interdigitated*. While elements of those perceived to have a higher prestige, namely politicians and mainstream media have diverted some attention to the Singapore blogosphere it has gained an element of *visibility*. The Singapore blogosphere does however seem to be showing

little interest in reaching out to control or gain action and so the flows of information remain within its own boundaries similar to that of an interface discipline (see table 1.). The quality of the blog posts and the quality of sources is the dominant valuation order of the Singapore blogosphere leading to the conclusion that it is an interface discipline. An interface discipline requires a level of commitment from the bloggers towards quality; this can be quality in a publication, fame, number of comments, or hyperlinks from others.

As Fisk, Cherney, Hornsey and Smith (2009:58) state using *Leximancer* to develop a country profile on core concepts is an ongoing process, “with profiles updated when newly discovered and emerging text sources are accessed”. This paper asserts that the Singapore blogosphere is not an isolated and distinct network without theme or focus. As fig.1 clearly shows, it is split along ethnic factions and has themes that arise, dissipate and re-occur over time. This ‘assigned normality’ of flux - contrasts with the fixed and rigid nature of the Lin, Sundaram, Chi, Tatemura, and Tseng’s, (2006) reading of the Singapore blogosphere.

REFERENCES

- [1] Atkinson, P. (1992). *Understanding Ethnographic Texts*. Sage Publications Newbury Park, Calif.
- [2] Borgatti, S.P. (2002). *NetDraw: Graph Visualization Software*. Harvard: Analytic Technologies.
- [3] Bruns, A. (2007). Methodologies for mapping the political blogosphere: An exploration using the IssueCrawler research tool. *First Monday*, volume 12, number 5 (May 2007), URL: http://www.firstmonday.org/issues/issue12_5/bruns/index.html.
- [4] Cavanagh, A. (2007). *Sociology in the Age of the Internet*. Open University Press.
- [5] de Nooy, W., Mrvar, A., & Batagelj, V. (2005). *Exploratory Social Network Analysis with Pajek*. Cambridge University Press.
- [6] Deleuze, G., & Guattari, F. (2004). *Anti-Oedipus: Capitalism and Schizophrenia*. Continuum International Publishing Group.
- [7] Dreyfus, H. L. (2001). *On the Internet*. Routledge.
- [8] Fisk, K., A. Cherney, M., Hornsey and A., Smith. (2009). "Rebuilding institutional legitimacy in post-conflict societies: an Asia Pacific case study, Phase 1A." University of Queensland. Accessed on 04/10/2009 at <http://espace.library.uq.edu.au/view/UQ:178786>.
- [9] Garrido, M. & Halavais, M. (2003). Mapping networks of support for the Zapatista movement. In McCaughey, M. & Ayers, M.D. (2003). *Cyberactivism: Online Activism in Theory and Practice*, Routledge.
- [10] Habermas, J. (2006). Political Communication in Media Society: Does Democracy Still Enjoy an Epistemic Dimension? The Impact of Normative Theory on Empirical Research I. *Communication Theory*, 16(4), 411-426.
- [11] Hine, C. (2005). *Virtual Methods: Issues In Social Research On The Internet*. Berg Publishers.
- [12] Howard, P. N. (2002). Network Ethnography and the Hypermedia Organization: New Media, New Organizations, New Methods. *New Media & Society*, 4(4), 550.
- [13] Hurst, M. W. (2006). Interactive Map of the Blogosphere. http://datamining.typepad.com/data_mining/2006/07/interactive_map.html.
- [14] Jackson, M. H. (1997). Assessing the Structure of Communication on the World Wide Web. *Journal of Computer-Mediated Communication*, 3(1).
- [15] Kendall, L. (1999). Recontextualizing “cyberspace”: Methodological considerations for online research. In S. Jones (Ed.), *Doing Internet research: Critical issues and methods for examining the net* (pp. 57 - 74). Thousand Oaks, CA: Sage.
- [16] Knox, H., Savage, M., & Harvey, P. (2006). Social networks and the study of relations: networks as method, metaphor and form. *Economy and Society*, 35(1), 113-140.
- [17] Lin, Y. R., Sundaram, H., Chi, Y., Tatemura, J., & Tseng, B. L. (2006). Blog Community Discovery and Evolution Based on Mutual Awareness Expansion. *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, 48-56.
- [18] Lyon, D. (1994). *The Electronic Eye: The Rise of Surveillance Society*. Polity Press.
- [19] Lyon, D., & Zureik, E. (1996). Surveillance, privacy, and the new technology. *Computers, surveillance, and privacy*, 1-18.
- [20] Mann, C., & Stewart, F. (2000). *Internet Communication and Qualitative Research: A Handbook for Researching Online*. Sage Publications Inc.
- [21] Markham, A. N. (1998). *Life Online: Researching Real Experience in Virtual Space*. Rowman Altamira.
- [22] McNally, R. (2005). Sociomics! Using the IssueCrawler to map, monitor and engage with the global proteomics research network. *Proteomics*, 5(12).
- [23] Miller, D., & Slater, D. (2000). *The Internet: An Ethnographic Approach*. New York University Press.
- [24] Misch, A. and White, H.C. (1998). Between Conversation and Situation: Public Switching Dynamics across Network Domains, *Social Research* 65:695-724.
- [25] Padgett, J., & McLean, P. (2006). Organizational Invention and Elite Transformation: The Birth of Partnership Systems in Renaissance Florence I. *American Journal of Sociology*, 111(5), 1463-1568.
- [26] Park, H. W. (2003). Hyperlink network analysis: A new method for the study of social structure on the web. *Connections*, 25(1), 49-61.
- [27] Radcliffe-Brown, A. R. (1940). On Social Structure. *Journal of the Royal Anthropological Society of Great Britain and Ireland*, 70(1-12).
- [28] Reid, E. (1996). Informed consent in the study of on-line communities: A reflection on the effects of computer-mediated social research. *The Information Society*, 12(2), 169-174.
- [29] Renninger, K. A., & Shumar, W. (2002). *Building Virtual Communities: Learning and Change in Cyberspace*. Cambridge University Press.
- [30] Roche, X. & et al. (n.d.). HTtrack 3.43-3 final.
- [31] Rogers, R. (2006). Mapping Web Space with the Issuecrawler. Online: <http://www.govcom.org/viewed/12.03.2008>.
- [32] Rutter, J. and Smith, GWH (2005) Ethnographic Presence in a Nebulous Setting, in *Virtual Methods: Issues in Social Research on the Internet*, edited by Hine, C., Berg, Oxford.
- [33] Scott, J. (2000). *Social Network Analysis: A Handbook*. Sage.
- [34] Siapera, E. (2006). Multiculturalism online: the internet and the dilemmas of multicultural politics. *European Journal of Cultural Studies*, 9(1), 5.
- [35] Smith, A., & Humphreys, M. (2006). Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping. *Behavior Research Methods*, 38(2), 262.
- [36] Stockwell, P., R. Colomb, et al. (2009). Use of an automatic content analysis tool: A technique for seeing both local and global scope. *International Journal of Human-Computer Studies* 67(5): 424-436.
- [37] White, H. (2008). *Identity and control: how social formations emerge*. Princeton University Press.
- [38] White, H., & Godart, F. (2007). Stories from Identity and Control. *Sociologica: Italian Journal of Sociology on line*, 3.
- [39] Wilson, B. (2006). Ethnography, the Internet, and youth culture: strategies for examining social resistance and “online-offline” relationships. *Canadian Journal of Education*, 29(1), 307.