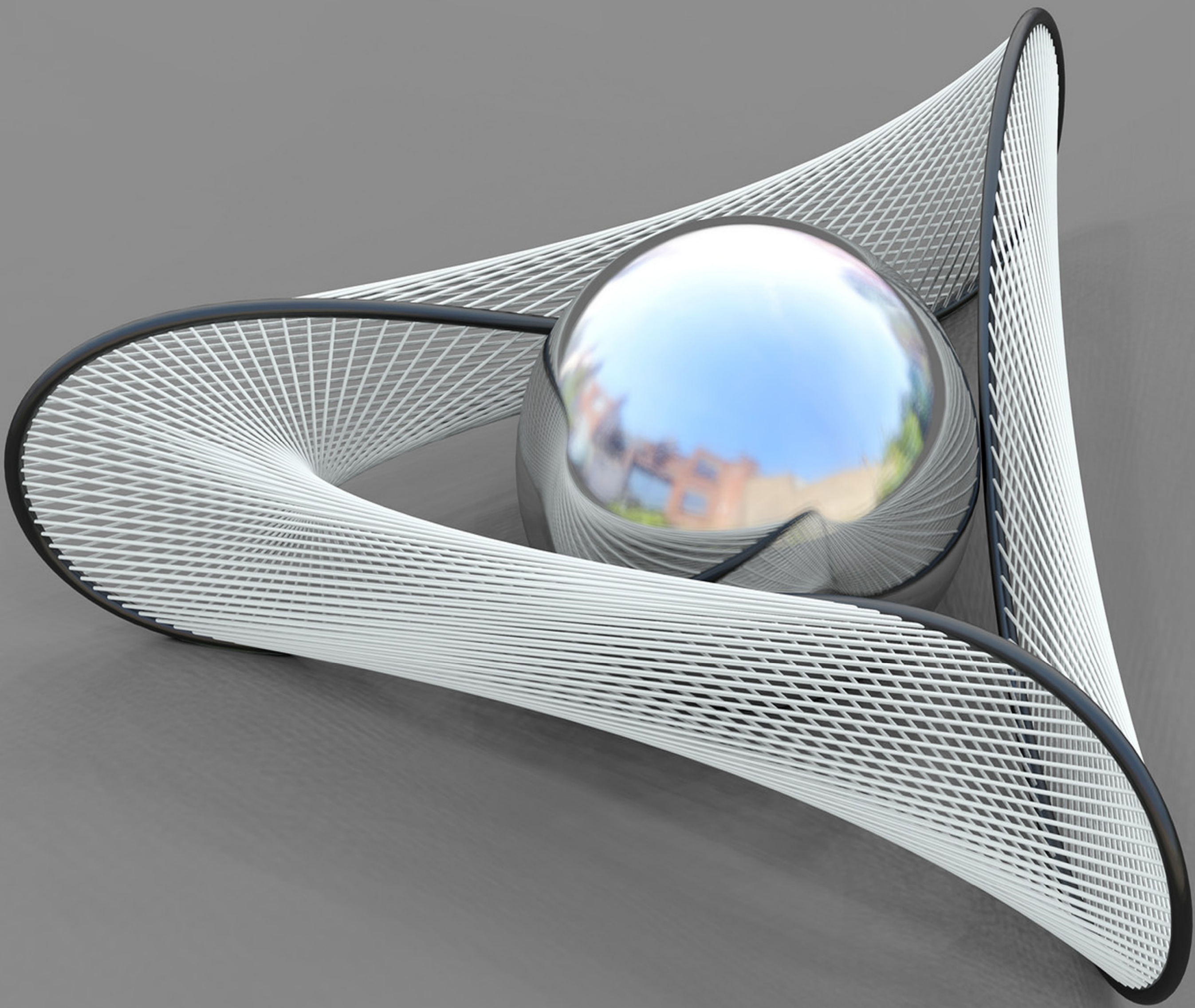
Journal of
Virtual Worlds Research
jvwresearch.org ISSN: 1941-8477

Assembled 2016

September 2016 (Part 2)
Volume 9 No. 2



Volume 9, Number 2
Assembled 2016 (Part 2)

September 2016

Cover image: Hypocycloid. Francesco De Comite, University of Lille
https://flic.kr/p/HCC6YR

Journal of
Virtual Worlds Research
jvwresearch.org ISSN: 1941-8477

Volume 9, Number 2
Assembled 2016 (2)
September, 2016

Detecting Covert Networks in Multilingual Groups: Evidence within a Virtual World

Janea Triplet
College of Southern Idaho, ID, USA

Andrew Harrison
Lindner College of Business
University of Cincinnati, OH, USA

Brian Mennecke†
Iowa State University, IA, USA

Akmal Mirsadikov
Iowa State University, IA, USA

† Brian Mennecke unexpectedly passed away on July, 9, 2016, during the revision of this manuscript. Brian is greatly missed by his family, friends, colleagues, and former students.

Abstract

This paper introduces an approach for the examination and organization of unstructured text to identify relationships between networks of individuals. This approach uses discourse analysis to identify information providers and recipients and determines the structure of covert organizations irrespective of the language that facilitate conversations between members. Then, this method applies social network analytics to determine the arrangement of a covert organization without any a priori knowledge of the network structure. This approach is tested and validated using communication data collected in a virtual world setting. Our analysis indicates that the proposed framework successfully detected the covert structure of three information networks, and their cliques, within an online gaming community during a simulation of a large-scale event.

1. Introduction

In an era when criminal and terrorist organizations have fully appropriated the Internet and social media to coordinate their activities and recruit followers, the need for new and adaptable approaches for analyzing communication activities has never been greater. The challenge is that criminal and terrorist networks are generally latent and often have no formalized organizational structure that is observable to an outsider (Ressler, 2006; Sageman, 2011), and often the accuracy of network models communication can only be tested in retrospect after the networks have been disassembled or destroyed (Carley et al., 2003B). We posit that a more practical approach deduces network structures as relationships evolve. A better understanding of the relationships within global covert networks is critical for researchers and practitioners to gain insights about how loosely affiliated international groups operate and can be observed, dissected, or disrupted (Ressler, 2006; Xu and Chen, 2008). Social network analyses allow scholars to describe these relationships using models of communication and dialogue (Moser et al., 2013). However, the process of making sense of communications patterns often relies on a preexisting understanding of the structures of organizations, and therefore analyses are seldom applied to evaluating covert networks (Xu and Chen, 2008). Analysts determining the structure of covert criminal or terrorist networks face three major challenges: (1) no formal network structure exists because the network is a series of individuals and smaller groups coupled together in various degrees (Carley et al., 2003B; Xu and Chen, 2008), (2) there are multiple languages and codes being used in the network (Malm et al., 2011), and (3) outside analysts cannot employ sophisticated tools based on semantics in real-time because they do not have a deep understanding of the languages or codes being used during communication (Davis, 2006).

In this paper, we use discourse analysis and social network analysis to deduce the latent structures of covert organizations in EVE Online, a massively multiplayer online game (MMOG) virtual environment. We use EVE Online as the context for our study because it represents a dynamic environment where large-scale events routinely take place as factions battle for territory, subvert opponent infrastructure, or create chaotic events. Another advantage of testing our model in a virtual world environment is that all of the communication can be captured, and all of the parties involved in communication can be identified in terms of formal and informal rank in the organization for the purposes of model validation (Bainbridge, 2007; Davis, 2009). This study also illustrates the utility of virtual world research in elucidating some nefarious behaviors that are difficult to accurately observe or measure in physical reality (Dilla et al., 2013).

Our research aims to answer the following research question: are there simple but effective means to discover the roles people play in covert organizations without clearly understanding the content of their messages? A secondary question examines the effectiveness of a virtual world environment as a platform for examining these phenomena. To answer these questions, our research develops a parsimonious approach for determining how to ascertain the structure of a covert network from observed communication. Specifically, we propose using discourse analysis to identify and extract semantic meaning from communications and then applying social network analysis to determine organizational structure. The advantage to this approach is that it determines the structure of the communication network irrespective of the technological artifact being used, the language of communication, and the context of communication.

The paper is organized as follows. First, we review the literature pertinent to network and discourse analysis in the context of understanding network structures and member behaviors. We follow this with a presentation of the methodology and research framework. The paper concludes with a discussion of the results and their implications for research and practice.

2. Background and Calculations

2.1 Detecting Structures in Criminal and Terrorist Networks

Modern covert networks often lack a formal leader and have an informal network structure with clusters of affiliated individuals (Ressler, 2006; Sageman, 2011). Having an obscured, and often unofficial, network structure makes it more difficult for outsiders to determine which actors play key roles within covert organizations (Shapiro, 2013; Dombroski et al., 2003). However, even when no official structure exists, covert networks still have actors who play critical roles in terms of coordination and management of resources (Carley et al., 2003A). For example, in a Global Salafi Jihad network, a few focal actors acted as links between the people who carried out the 9/11 attacks and the 2002 terrorist bombings in Bali, and on average, each member of the network was only 2.5 links away from Osama bin Laden (Xu and Chen, 2008). Similarly, international drug trafficking enterprises may have loosely clustered groups of suppliers collaborating with highly organized groups, involved in large-scale shipment, and may utilize independent individuals for direct distribution (Malm and Bichler, 2011). Despite often having ambiguous official structures, control and coordination mechanisms do exist within covert networks and these structures can be modeled to identify key actors (Carley et al., 2003B).

In addition, global covert networks often span national, cultural, ethnic, and linguistic borders (Malm et al., 2011) making single-language analytical strategies ineffective. Military and law enforcement agencies have been unable to acquire an ample number of language specialists that are capable of deciphering messages in networks, using multiple languages (Cordesman, 2003; Davis, 2006; Benjamin, 2007). Accordingly, analytical techniques that assume an in-depth semantic understanding of the content of each message may be realistic in some settings, but are untenable when evaluating covert network structures. Virtual world environments possess similar covert structures and, as a result, may represent useful platforms for testing theories and analytical techniques that deduce these covert structures (Bainbridge, 2007).

2.2 Using Discourse Analysis

Discourse analysis is a useful method for identifying the key features and patterns within messages within groups linked by common goals, communication, and values (Abdul-Gader & Kozar, 1990; Paltridge, 1998; Hymes, 1972; Bhatia, 2004; Labov & Fanshel, 1977; Milroy, 1987; Swales, 1990). Discourse analysis has been applied to study organizational boundaries and control structures based on communications (Alvesson & Karreman, 2000). These methods have proved particularly useful for the relationship between communication and power among group members (Kalou and Sadler-Smith, 2015). Most extant discourse analyses focus on interpretive approaches that require a deep understanding of social, psychological, and semantic contexts of communication (Alvesson & Karreman, 2000). These interpretive approaches are effective for deducing social relations when there are no constraints limiting the effort, money, or time required to conduct such analyses. Consequently, the majority of discourse analyses have been applied in contexts where communications are obvious and interpretable (Paltridge, 2012).

The timely interpretation of the meaning of words is difficult when analyzing covert networks (Carley et al., 2003B). However, other symbolic components of communication are available. Covert networks are united by either a shared ideology or a shared procedural goal, such as accruing money via drug or weapon sales or carrying out a terrorist attack (Dombroski, 2003; Shapiro, 2013). Even in loosely coupled covert organizations, communication routines emerge as a means to communicate efficiently and clearly to achieve shared goals (Paltridge & Burton, 2000). Thus, discourse analysis offers an appropriate analytical lens for identifying which key communicative features are consistent within an international covert network (Johnstone, 2000; Paltridge, 2012). For understanding communication patterns, some of the most useful features of communication are indicators of requests for information (Kwong & Yorke-Smith, 2012; Cao et al., 2011). The approach to analyzing covert networks that is introduced in this paper, works backwards by evaluating communications patterns using discourse analysis to identify simple, language-neutral keys to information seeking requests. Once we identify cues in communications that indicate requests for information, we then use these requests as a means to study the social interactions and latent hierarchical structure of the network.

2.3 Features of Covert Social Networks

Social network analysis is a powerful analytical method that can be paired with discourse analysis (Granovetter, 1973; Goyal et al., 2006). However, difficulties persist in applying social network analysis to study covert organizations, where the extent and structure of networks are obscured (Dombroski et al., 2003; Carley et al., 2003B; Ressler, 2006). Covert networks share similarities with traditional networks in terms of the efficient flow of information and resources, but also differ in some regards. Covert networks tend to have high clustering, short average path lengths (Xu and Chen, 2008), and are vulnerable to disruption in key gap spanners (Carley et al., 2003A; Shapiro, 2013). International covert networks exacerbate this characteristic as a few key individuals play the role of translators across subgroups and cliques (Malm et al., 2011).

In addition to challenges related to interpreting the meaning of messages, there are methodological challenges associated with applying social network analysis to covert networks (Carley et al., 2007). Many of the complex analytical strategies for evaluating covert networks assume a total collection of messages and a complete understanding of their content. For example, defining question answer pairs using words like “what”, “when”, “where”, “why”, and “how” uses a more accurate algorithm (Kwong and Yorke-Smith, 2012), but at the cost of training an analyst in language skills necessary to identify and interpret the content of every message in the language in which the message was written. Therefore, social network analyses of covert criminal networks are almost always entirely used post hoc, at which point many actors have already been captured or killed (Dombroski et al., 2003; Xu and Chen, 2008; Ressler, 2006; Mam and Bichler, 2011). A greater value exists in identifying and disrupting criminal covert networks in near real-time during operations. Consequently, our focus is on using simplistic social network analysis strategies applied, in near real-time, without any a priori knowledge of the network structure or a deep understanding of the contents of each message.

Covert networks often also feature dynamism and individuals’ roles may change as network members are captured, killed, or go into hiding (Carpenter et al., 2002). As events unfold, some agents will become more or less central to action and communications efforts (Magouirk et al., 2008). These changes to network structure can undermine traditional static measures of centrality (Carley et al., 2007). As a result, additional measures of network structures including betweenness centrality, out-degree/in-degree relations, and cluster coefficients, make analyses more robust to dynamism. Dynamic criminal and terrorist networks are also often decentralized (Sageman, 2011). Organized crime and terrorist networks often have clusters at different levels of analysis due to features such as kinship, where smaller criminal cells operate independently within larger organizations (Magouirk et al., 2008; Tsvetovat & Carley, 2007). Resultantly, it is critical to analyze the organization at various levels of analysis to model decentralized structures. Finally, identifying key peripheral relationships is particularly useful in the context of evaluating the structure of covert criminal and terrorist networks. It may be necessary to rapidly model organizational networks during extraordinary events and crises to determine if other agents within a network pose similar threats (Pfeffer & Carley, 2012). Consequently, an automated, or semi-automated process for analyzing network structure can be executed in a more timely fashion.

3. Methodology

3.1 Research Setting

We collected data from a virtual world setting to empirically test the usefulness of our discourse analysis-based algorithm. The virtual world environment selected for the study was EVE Online, an international MMOG. EVE Online contains critical features of robust virtual worlds including an active community, the creation of goods, and commerce (Sivan, 2008). Specifically, we selected this setting because it contained complex communication behaviors by a number of loosely affiliated subgroups, was event-driven, and exhibited a high degree of environmental uncertainty. Individuals within the network collaborated towards a shared goal, used multiple languages, coded phrases, and jargon, developed their own norms, obscured their true identities, and orchestrated resources and roles within the organization.

We selected this research environment to collect data because it parallels many of the environmental characteristics of real-life covert and criminal networks. Behaviors in virtual worlds largely mimic the behaviors of individuals in the real-world (Herring (2004)) and offer researchers a manner to formally study natural, real-time, behaviors that may not be readily observable in the real world (Davis, 2009). Virtual worlds offer rich social interaction and collaboration that parallel real-world communications in many facets (Messinger et al., 2009). Furthermore, virtual gaming environments are renowned for their dynamic and unpredictable nature (Bainbridge, 2007). However, in a virtual world, participants were not actually engaged in crimes or war allowing researchers and participants to freely discuss their behaviors.

After obtaining Institutional Review Board approval, we examined communications occurring during a series of events in which hundreds of participants collaborated to claim and attack property (i.e., virtual trade networks) affiliated with a different coalition of people. The goals of this type of event typically involve the disruption of virtual trade networks of rival organizations, and the acquisition of virtual goods and territory. These events can take days, weeks, or months to occur and require participation and cooperation to coordinate the actions of hundreds of individuals with varying degrees of allegiance. This research setting provided a vibrant and culturally rich environment where communication patterns and organizational structures were developed and maintained by the community of participants. For example, in EVE Online players may align themselves with one of five major organizations, or may join specialized self-governed alliances. Each of the competing alliances jostle for resources and power and are comprised of thousands of loosely affiliated groups of individuals who ally to accomplish shared goals. These groupings represent important social institutions that facilitate collaboration for achieving various goals. EVE Online also contains an extensive archive of discussion threads, forums, and community pages. The game 'universe' encompasses character profiles, wikis, and a detailed encyclopedia of gaming jargon (i.e. words, phrases, acronyms and emoticons known and used by this discourse community). The history of the game was chronicled in player created backstories, art, and video. Demonstrating the global involvement of this MMOG, the player guide was available in English, German, and Russian.

3.2 Data Collection

We began by collecting a large set of unstructured text communications and applying discourse analysis to recognize patterns and structures within any messages. The study enlisted an informant for data collection.[1] Ethnographers use informants to gain access to field sites with barriers of entry to outsiders (Lofland et al., 2006). A double-blind data collection approach was used to avoid convoluting the natural evolution of events and to reduce researcher bias in the discourse and network analyses. The informant who recorded communication was unaware of the study's purpose. The researchers were uninformed as to the organizational structure of the MMOG, the identity of the players, or the purpose of communication. The researchers did not have direct, unobservable, entry into to the virtual community or a pre-existing membership within any organizations. During the data collection, the informant recorded chat activities from Internet Relay Chat (IRC) channels for three different virtual communities. IRC is an open communication system designed for real-time text messaging over the Internet. The system is primarily used for group communication through discussion forums called 'channels.'

The data were collected over 14 days in 2011, and were collected for multiple groups allowing analysis of small, medium, and large group communication. Having three levels of analysis (i.e., small, medium, and large groups) provided different views of the same phenomena and each of these different views may each offer unique insights into communications. The first group (i.e. the small group) was comprised of 10 participants who wrote 798 lines of text and 4,149 words. The small group represented the use of a two-column edge list to describe information seeking and information providing behaviors between small numbers of actors. Analysis at the level of small groups was useful in interpreting how a tightly-focused, task-specific, topic was shared between individuals in a conversational tone. Thus, small networks are useful in interpreting social interaction and group conventions. The second group (i.e. the medium group) was represented by 325 participants who wrote 1,319 lines of text and 4,644 words. The medium group level of analysis added context about the unequal distribution of power and how perceptions of authority or expertise may affect communication efforts. The third and final group (i.e. the large group) was made up of 683 participants who produced 4,178 lines of text and 21,090 words. The large group analysis was useful for understanding clustering and bottlenecks in the communication network.

3.3 Interpretive Framework

We utilized the interpretive framework in Figure 1 to manage the large amount of unstructured text collected for the analysis. Once the data were collected, we used a semi-automated formula to reduce coding error and decrease the amount of time needed to create an edge-list for social network analysis. First, the chat logs were imported into a spreadsheet. Three attributes were used for analysis: (1) the date and time of the communication event, (2) participant's gaming name and (3) any messages written by the participant. The formula we developed to analyze the messages was based on discourse analysis. Discourse analysis allows us to understand how individuals use messages to coordinate and interact with each other during events (Johnstone, 2008; Paltridge, 1998). The analysis of discourse created by MMOG players employed speech act principles such as turn-taking (Sacks et al., 1974),

[1] The study was undertaken with approval of the human subjects review board of the researchers' institution.

conversational contribution (Grice, 1975), directives (Gordon & Lakoff, 1971), and requests (Labov & Fanshel, 1977). Communicative interactions examined under the lens of discourse analysis offered insight into status, socialization, rights, obligations, needs, and abilities of the virtual community. Similarly, Moser and colleagues (2013) had combined discourse analysis and social network analysis to examine networks in an online occupational community. Further, discourse network analysis (DNA) (Leifeld & Naunss, 2012; Fisher, Waggie, & Leifeld, 2013) has been proposed as a means to examine political networks. This methodology involves a combination of content analysis and social network analysis by looking for affiliation, concept congruence, actor congruence, actor conflicts, and dynamic forms of these discourse networks.

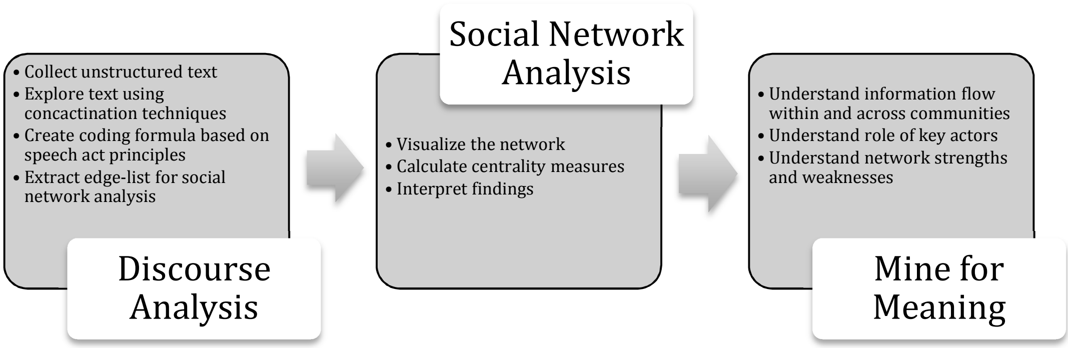


Figure 1: Interpretive Framework

To explore the structure of communication, concordance techniques were used (i.e. word frequency, key-word-in-context). A theme of information "seeking and providing" emerged from this process. The text chat occurred in sequential order, as is consistent with the observation of turn-taking in other discourse analyses (Van Dijk, 1993; Potter, 1996; Paltridge & Burton, 2000). When an individual is asked a question, people interested in the question, or obliged to respond to that individual, will quickly respond (Belkin et al., 1987). Consequently, we utilized this pattern of knowledge elicitation to determine how information was being distributed and processed within the network. Importantly, the theme of seeking and receiving information was observed in both English and Russian chat, and represented a language-independent semantics revealed by discourse analysis. The players used specialized punctuation (i.e. a question mark) to signify the end of questions for both English and Russian chat. We expect that punctuation is a reflexive component of information seeking communication behaviors, we also consider that it might be left off of some messages. However, we intentionally prioritized simplicity and selected a naïve algorithm where an extensive understanding of language translation was unnecessary.

Punctuation patterns were the inspiration for our development of a parsimonious function that forms the core of the analysis technique we used to discover the network structures. The function searched the text for question marks, then selected that row (i.e. ego) and the next three rows (i.e. possible alters), and moved those rows to a separate worksheet. Question and answer pairing has been shown to be an effective means to construct a social network structure for e-mail based communication

(Kwong and Yorke-Smith, 2012) and Internet forum posts (Alvarez et al., 2010; Hong and Davidson, 2009; Sun et al., 2010).

The decision to associate three responses with each question stemmed from two key themes: (1) the number of responses before a question was answered within a random sample of data, and (2) qualitative evidence from a participant that almost all questions are answered almost immediately. The random sample consisted of 10% of the data from each edge list, and was analyzed using the following strategy. Questions were identified from the text during discourse analysis. Then, the number of responses before the question was answered were counted. Of the questions asked in the data subset, 95% were answered within the first three responses following the question. After the data had been collected, a participating informant in the study had also indicated that elicited responses are nearly immediate and validated the number of responses.

3.4 Social Network Analysis

An open source software application, NodeXL (Smith et al., 2010), was used to perform the social network analysis to understand the relative importance of individual roles within the emergent operational hierarchies. The social network analysis illustrated relevant patterns of communication activities (i.e. information gathering and providing). The application makes use of a two-column edge list to explore the relationships (i.e. edges) between actors (i.e. nodes). We used the discourse analysis to inform the structure of the edge-list based on the theme of information seeking and providing. Human interpretation of the text for meaning was unnecessary because the function was built on the larger act of asking a question and not on specific words. The network structure was built using inferred information seeking and providing relationships between egos and agents. The Fruchterman-Reingold layout algorithm was used to visualize the network.

Social networking metrics were calculated for: eigenvector centrality, betweenness centrality, clustering coefficient, closeness centrality, vertex degree, and in-degree/out-degree. Eigenvector centrality measured the importance, prestige, prominence, and power of an actor in a network (Freeman, 1979). Eigenvector centrality was calculated using the accelerated power method outlined by Borgatti (1995). Eigenvector centrality measures how well an actor is connected within the entire network, and not just within their local relationships. Actors with high eigenvector centrality represent leaders within the network.

In comparison, betweenness centrality measures the shortest distance of one actor to others in the network and represents indicates which individuals have influence over information flow within the group (Brandes, 2001). As a result, individuals with high betweenness centrality are brokers of information because while they may not have the greatest number of shortest paths to other individuals, the shortest paths of the most other individuals frequently go through them. Thus, both measures are important for understanding how information flows throughout an organization, because eigenvector centrality evaluates the shortest possible routes for information to flow through, while betweenness centrality measures the importance of a single individual in controlling the flow of information. In the directed graph, the out-degree reports the number of outgoing edges (i.e., the number of questions asked by an individual). The in-degree reports the number of incoming edges (i.e., the number of actors responding to questions). The cluster measures how close the actor and its neighbors are to forming a clique or sub-group. In social groups, actors with high degrees of betweenness centrality play powerful roles in the network and often have control over the flow of information (Newman, 2004).

The clustering coefficient of a vertex in a graph is determined with a formula that quantifies how close the vertex and its neighbors are to being a clique. It offers an alternative metric that indicates power roles within groups – it can identify individuals who may not communicate often but hold great sway over the behaviors of others.

While it is important to identify the existence of sub-groups or cliques, it is also important to find the individuals who frequently connected their clique to other sub-groups through communication exchanges. We evaluated closeness centrality to find the communication connectors or the boundary spanners within the network. Boundary spanners often serve a powerful role as a network connector and yet this measure could also indicate possible weaknesses within the group's communication flow. If a connector was absent from the group, then the information flow between certain cliques may fail.

4. Results

4.1 Small Group Discourse and Social Network Analyses

In the small group IRC channel, ten players produced 798 communication exchanges over 14 days. Three individuals contributed 77 percent of the conversation. The conversation switched between Russian (57 percent) and English (40 percent). The group also exchanged non-verbal communication codes (3 percent) through various emoticons (e.g., :), :p, :(, o/) and internet acronyms (e.g., lol, kk, afk, gn, TS, noob). The content of the online chat included conversations about tasks, internal politics, character assessments, relationship building, advice, complaining, banter, and humor. No chat message was over 10 words long. Longer messages were split into shorter phrases and entered consecutively by a player who had more to say. The group used abbreviations, acronyms, and game-specific jargon.

A network diagram created from the discourse pattern (Figure 2) indicated that individual S01 was the center hub of information exchange. This member was directly linked to all but one of the group members. Three of the remaining nine members of the group only spoke directly to S01. The second connector in the group was individual S07, who was directly linked to four other players. Individual S07 had the most lines of chat, but communicated with a smaller number of people. Group member S06 connected an outlier (S08) to the rest of the network. The network diagram illustrated that there were two subgroups or cliques within this network. Individual S01 was the hub of the group and connected three outliers (participants S02, S03, and S06) to the others within the group. The connection between individuals S01 and S06 spanned a boundary and appeared to connect this group to another group represented by individual S08.

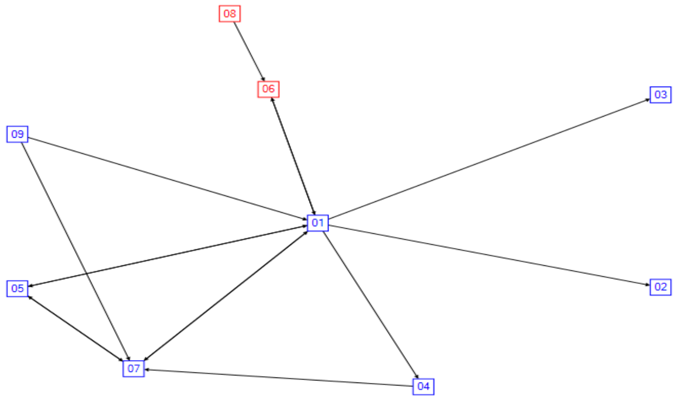


Figure 2: Network Diagram of Small Group Interaction

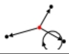
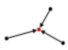
4.2 Medium Group Discourse and Social Network Analyses

The medium group exchanged 1,319 messages consisting of 4,644 words among 325 participants. One individual contributed 264 lines of chat text and 135 individuals contributed one line of text chat each. Similar to the small group communication pattern, the medium group was depicted by a power-law distribution. The power-law distribution in social networking illustrates the voluntary nature of the environment (Shirky, 2008). When participation is voluntary, the long-tail indicates that participants are free to join the group and contribute when motivated to do so.

A network diagram was created representing type of communication exchanges produced by the medium size group (n=325). After running the information answering-seeking formula, a sample of 149 communication exchanges was used to construct the edge list for the purposes of social network analysis. What was determined from the directional graph was that there were three types of communication roles within the group. Participants were information seekers, information providers, or a mix of the two. The role of information seeker was represented 55 percent (n=82) of the communication exchange. The role of information provider represented 29 (n=43) percent of the communication exchange. Participants whose text chat was a mix of both information provider and seeker represented 16 percent (n=24) of the total communication exchange (n=149).

Within the medium-sized group, individual M01 was the top contributor of text by producing 20 percent (n=264) of the total communication exchanges (n=1,319). Individual M01 was also the top information provider of the group. Individual M158 was the group's top information seeker. Participants M02, M08, and M09 produced communication that was a mix between information provider and seeker. As shown in Table 1, the network was sorted by betweenness centrality. Nodes with higher betweenness centrality measures indicate important locations within the group by showing the individuals who have influence over the information flow within the group (Brandes, 2001). Individual M01 was the most frequent communicator and produced 264 lines of text. Individual M01 also received the highest measure of betweenness centrality. The individual with the second highest betweenness centrality ranking, however, was individual M09, who was the ninth most frequent communicator producing 16 lines of text. Next, individual M158 received the fourth highest betweenness centrality ranking and, yet, produced only two lines of text. As the sub-graph for individual M158 indicates, a well formed star pattern of information exchanges suggesting many linkages to others within the group.

Table 1: Comparison of Betweenness Centrality Ranking versus Lines of Text Produced

Medium Group Participant ID	Lines of Text Produced	Subgraph	In-Degree	Out-Degree	Betweenness Centrality
M01	264		25	0	1.000
M09	16		3	8	0.316
M39	6		0	3	0.302
M158	2		0	5	0.278
M11	13		3	0	0.274

4.3 Large Group Discourse and Social Network Analyses

In the large group, there were 4,178 communication exchanges between 683 participants consisting of 21,090 words. Twelve individuals contributed 624 lines of chat text and 175 participants contributed one line of text each. We used the same strategy as the medium group to create an edge list using our information-seeking algorithm. The data listed in Table 2 indicates that there are players who were not top communication contributors but who were high on the betweenness centrality ranking. People with high betweenness centrality and a low volume of communication, do not contribute their own information, but often control the flow of information within a network (Brandes, 2001). For example, individual L22 was the 10th most talkative participant by producing 28 lines of text and was ranked 20th on the betweenness centrality measure. Conversely, player L08 was the 15th most talkative player by producing 19 lines of text and was ranked 6th on the betweenness centrality measure.

Table 2: Social Network Analysis Summary for the Large Group.

Eigenvector Centrality		Betweenness Centrality		OutDegree (Question)		InDegree (Answer)		Cluster	
Score	Large Group Participant ID	Score	Large Group Participant ID	Score	Large Group Participant ID	Score	Large Group Participant ID	Score	Large Group Participant ID
0.236	L09	1.000	L04	25	L09	27	L04	1080	L09
0.219	L06	0.878	L06	22	L06	26	L02	1080	L14
0.201	L05	0.700	L05	18	L05	20	L14	1080	L17
0.191	L02	0.631	L02	16	L08	20	L10	1077	L02
0.188	L15	0.594	L15	15	L27	18	L01	1077	L10
0.171	L27	0.557	L08	15	L15	17	L05	1077	L06
0.159	L17	0.523	L14	14	L24	14	L20	1077	L12
0.158	L08	0.504	L09	13	L23	14	L17	1077	L21
0.157	L01	0.430	L10	12	L26	13	L06	1070	L01
0.152	L04	0.382	L21	11	L25	12	L12	1070	L22
0.145	L14	0.339	L01	11	L21	11	L15	1070	L08
0.145	L21	0.300	L27	8	L01	10	L22	1070	L23
0.144	L20	0.265	L17	6	L17	9	L25	1069	L04
0.135	L25	0.261	L26	3	L02	9	L08	1049	L05
0.131	L23	0.237	L24	3	L22	8	L24	1049	L24
0.129	L10	0.237	L23	3	L04	8	L26	1049	L25
0.126	L12	0.192	L25	2	L12	6	L09	1049	L26
0.123	L26	0.150	L20	0	L14	6	L27	1014	L27
0.115	L24	0.109	L12	0	L20	6	L21	960	L20
0.111	L22	0.100	L22	0	L10	3	L23	960	L15

In this network, 190 individuals were assigned to 17 cliques, with the 20 top contributors representing seven of the 17 cliques. This indicates that power roles within the network were unevenly distributed. However, in cases where top contributors belonged to the same clique (e.g. 347), the purpose of communication often differed. For example, individuals L09 and L20 were members of clique 347. Individual L09 contributed 25 questions and 6 answers while individual L20 contributed 16 answers and no questions. These two were both talkative players from the same clique, but L09 was an information seeker and L20 was an information provider.

Assessments of closeness centrality displayed in Table 3 identified individuals who acted as boundary spanners between these cliques. Closeness centrality is the natural distance between a pair based on the shortest path of communication, and can indicate a more meaningful measure of power in an information-seeking network (Stephenson and Zelen, 1989). Individual L09 was the top boundary spanner with the lowest closeness centrality measure. This participant's communication-reach extended to 31 other individuals and to 10 other cliques. Similarly, Individuals L20 and L19 were strongly connected within their own cliques, but their exchanges with outside cliques were usually limited to one or two other individuals.

Table 3: Connections to Other Players and to Other Cliques

Boundary Spanners			Clique Assignment with Number of Connections to Other Cliques																Total Reach	
Own Clique	Large Group ID No.	Closeness Centrality	282	310	315	322	338	341	347	350	355	356	357	358	359	361	362	363	Other Players	Other Cliques
347	L09	2.370			3	1		2	13	3		2		1	1	1	2	2	31	10
315	L02	2.429			6		2		2	2	3	1	1		1	1		1	20	9
350	L17	2.471	1	1		1	2		1	5		1	1	2		2	3		20	10
350	L27	2.503		2	1		1	3	1	7		2	1				1	2	21	9
347	L15	2.519		1	1		1	1	5	1		2	3		1	1		1	18	10
341	L06	2.519				1	2	7	3	2	1	1	2			1			20	8
361	L12	2.556		1	1			2	1	2	1		2	1		3			14	8
350	L23	2.571			1			1		8		2	1		2	1			16	6
347	L20	2.640		2			2	2	4	1		1	1					1	14	7
310	L19	2.720		6		1	1		2			1		1		1	1		14	7

Figure 3 also illustrates that some cliques are better connected than others and that some cliques are more fragile than others. For example, clique No. 266 consisted of four members. Two of the members were connected only to each other. The other two were connected to their own clique and were separated from the top 10 boundary spanners by two degrees (e.g. L40—L38—L02 and L41—L42—L19). The information flow to and from clique No. 266 might be considered fragile. Consequently, this clique would have a difficult time communicating with almost every other individual if their access to a few other individuals was disrupted.

Interestingly, the size of the clique did not always determine its vulnerability to being cut off from the communication flow. Clique No. 322 also consisted of four members but three of the four were directly connected to the top 10 boundary spanners (e.g. L43—L09 + L17; L44—L06; and L45—L19). These communication lines illustrate that this sub-group had direct access to the top boundary spanners in the large network.

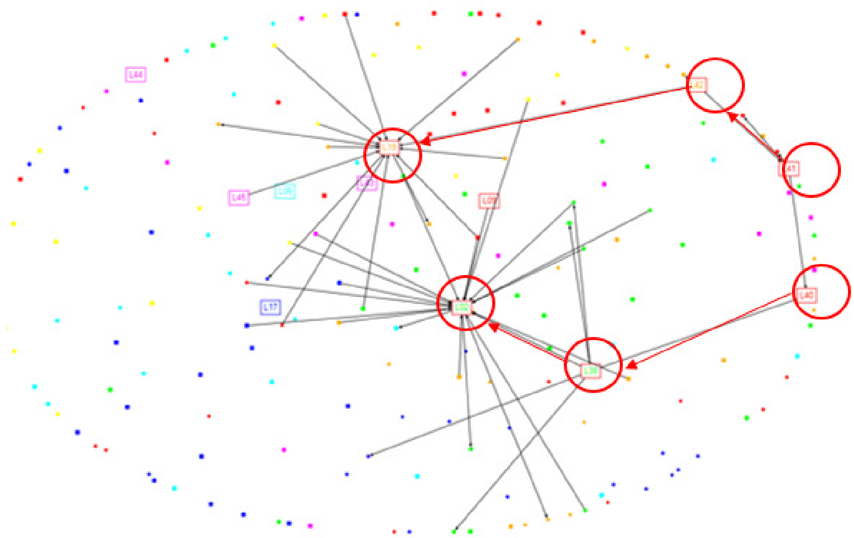


Figure 3: Clique No. 266 Illustrated Two-degrees of Communication Separation

4.4 Informant Validation of Analyses

Our data was collected during a series of conflicts between two large Eve Online alliances including a group subsequently referred to as the "Red Alliance", and another group referred to as the "Northern Coalition" (Drain, 2010). These virtual environments provide an opportunity to validate research models that may be otherwise untestable due to legal, social, or technical constraints (Davis et al., 2009). Thus, following our data collection and analyses we asked our informant to evaluate our findings in post hoc assessments. These post hoc assessments compared the networks deduced by our algorithms to the observable governance structures of the players. To validate our analysis of the small group structures, the informant was presented the network diagram and our expectations that S01 held a central leadership role and that player S07 held a secondary leadership position. When presented the player account names, the informant confirmed that S01 was the CEO of that corporation, and that S07 was the Director of the corporation. The CEO and Director are the two topmost positions in an EVE Online corporation, and indicate that our algorithm correctly identified the small group's leadership structure. Furthermore, the informant described that S08, was not part of the corporation, but was a representative of an allied group of players, as predicted by the small group analysis.

For validating the medium and large group structures, the informant was presented a list of players ranked by Eigenvector centrality and betweenness centrality, and the players' clusters as deduced by our algorithm. When presented the clusters using player names, the informant indicated that about 85% of the players the informant could identify were clustered within groups that corresponded to their EVE Online corporations and alliances. Furthermore, the informant estimated that the players with the highest eigenvector and cluster centrality rankings (e.g., L09, L02, L06, and L05) represented key members, referred to as fleet commanders, of an alliance. Likewise, the informant indicated that players acting as boundary spanners (e.g., L19 and L20), were relaying orders to smaller groups of players and translating commands from Russian to English. Consequently, the informant provided evidence that the clusters and leadership structures identified by our algorithm predicted the players' actual governance structures.

5. Discussion and Conclusion

The purpose of the study was to explore whether meaningful and accurate communication patterns could be detected using a simplistic strategy to identify question and answer pairs. Indicating the success of the study, the participating informant validated that the study accurately detected the actual network connections and sub-cliques as structured in the organization's para-military structure. The respondent indicated that these were unexpected for two reasons: (1) the communication was in a foreign language that should be unintelligible to the researcher, and (2) that much of the communication used codes, jargon, and acronyms. Thus, these observations provide support for the view that a simplistic language-independent heuristic can be used by non-experts to quickly identify key players in covert networks. This finding inspires confidence that despite having thinly-stretched resources spread across the globe, the judicious use of analytical strategies can be used to reveal covert global crime and terror networks. The proposed framework was also able to accurately illustrate a visualization that inferred group leadership and communication channels. Finally, we expect that studying virtual community networks, will present a useful avenue for future research to improve our capabilities to model covert organizational structures.

First, one of the interesting implications of this study is that it supports the methodological position that virtual worlds represent a useful and valid avenue for studying real time communication (Davis, 2009). While virtual worlds differ in certain social or spatial dimensions from real life, virtual world interactions provide insights into behaviors obscured in other settings (Bainbridge, 2007). Virtual worlds may be particularly useful when studying behaviors considered immoral or illegal in other contexts (Dilla et al., 2013). Furthermore, within the virtual world setting observation is non-intrusive. Data was collected based on real time communication logs and messages were neither disrupted nor distorted by observation.

Additionally, our research supported the use of punctuation marks as indicators of network relationships. The question mark has been examined in the context of forum and email communication as an indicator of seeking and response behaviors (Ding et al., 2008; Kwong & Yorke-Smith, 2012; Cao et al., 2011); yet, we are not aware of any research that has examined this method in the type of large-scale social network analysis of dynamic and interactive gaming environments. This strategy represents a useful tactic for developing accurate models of information sharing, because information seekers will generally ask others to provide useful details. We witnessed punctuation marks being used in both Russian and English communications, which suggests that the use of indicators separating questions from statements occurs across languages. As such, an important contribution of this research is to demonstrate that we can detect covert network structures using simplistic heuristics for detecting questions in environments that have similarities to other networks of interest (e.g., terrorist or criminal networks). Thus, this model is robust to changes in languages and requires little decoding or translation prior to application in determining covert network structures.

Furthermore, the data provided theoretical insights into the structure of covert networks. In summary, actors with high degrees of betweenness centrality played powerful roles in the network and exhibited control over the flow of information. The quantity of messages influenced information flow, but the quality and specific type of information also influenced a player's power in the group and their control over information. Our data also indicated that the fragility of communications within the network is not restricted to small subgroups. If access through a few key communication nodes is disrupted, larger groups of individuals can also be disconnected from the rest of the organization. These findings could have significant importance for both theory and practice.

Finally, our findings had practical implications for how social networks are vulnerable to disruption. Players who also exhibited frequent communication exchanges with other sub-groups were found to be boundary spanners within the network as indicated by the closeness centrality measure. These players held powerful influence over the flow of information within their own groups and throughout the network. This finding has strong implications for examining focal actors in criminal and terrorist networks. In particular, these findings indicate that certain individuals act as communications brokers and strongly influence the actions of people around them. Thus, our findings indicate that individuals within a criminal or terrorist communications network, who exhibit high betweenness and closeness centrality measures, are likely to be powerful figures in the functional operations of the organization, even if they hold no formal role. We also find that individuals that relay messages to subgroups constitute critical points in the communication network and may be where the most serious threats of information overload and information loss may occur. Consequently, without secondary channels of communication between subgroups to alleviate information bottlenecks and backup communications when main channels are lost or destroyed, individuals within large networks can quickly become isolated from the rest of the network.

References

Abdul-Gader, A.H., & Kozar, K.A. (1990). Discourse analysis for knowledge acquisition: the coherence method. *Journal of Management Information Systems*, *6*(4), 61-82.

Alvarez, H., Ríos, S. A., Aguilera, F., Merlo, E., & Guerrero, L. (2010). Enhancing social network analysis with a concept-based text mining approach to discover key members on a virtual community of practice. In R. Setchi, I. Jordanov, R.J. Howlett, L.C. Jain (eds.), *Knowledge-Based and Intelligent Information and Engineering Systems,* 591-600. Berlin: Springer.

Alvesson, M., & Karreman, D. (2000). Varieties of discourse: On the study of organizations through discourse analysis. *Human Relations*, *53*(9), 1125-1149.

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of twitter data. *Proceedings of the Workshop on Languages in Social Media,* 30-38.

Bainbridge, W. S. (2007). The scientific research potential of virtual worlds. *Science*, *317*(5837), 472-476.

Belkin, N. J., Brooks, H. M., & Daniels, P. J. (1987). Knowledge elicitation using discourse analysis. *International Journal of Man-Machine Studies, 27*(2), 127-144.

Benjamin, S. (2007, June 8). Don't ask, don't translate. *New York Times*, Retrieved from http://www.nytimes.com/2007/06/08/opinion/08benjamin.html

Bhatia, V. K. (2004). *Worlds of Written Discourse: A Genre-Based View*. London: Continuum.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, *2*, 113-120.

Borgatti, S. P. (1995). Centrality and AIDS, *Connections*, *18*(1), 112-115.

Brandes, U. A (2001). Faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, *25*(2), 163-177.

Carley, K. M., Reminga, J., & Kamneva, N. (2003A). Destabilizing terrorist networks. *Proceedings of the North American Association for Computational Social and Organizational Science Conference 2003,* Pittsburgh, PA.

Carley, K. M., Dombroski, M., Tsvetovat, M., Reminga, J., & Kamneva, N. (2003B). Destabilizing dynamic covert networks. *Proceedings of the 8th International Command and Control Research and Technology Symposium,* National Defense War College, Washington, DC.

Carley, K. M., Diesner, J., Reminga, J., & Tsvetovat, M. (2007). Toward an interoperable dynamic network analysis toolkit. *Decision Support Systems*, *43*(4), 1324-1347.

Carpenter, T., Karakostas, G., & Shallcross, D. (2002). Practical issues and algorithms for analyzing terrorist networks. Invited paper: *Western Multi-Conference*, 1-6.

Cao, Y., Yang, W. Y., Lin, C. Y., & Yu, Y. (2011). A structural support vector method for extracting contexts and answers of questions from online forums. *Information Processing & Management*, *47*(6), 886-898. doi: http://dx.doi.org/10.1016/j.ipm.2010.06.004

Cordesman, A. H. (2003). *The Current Military Situation in Iraq.* Washington, DC: Center for Strategic and International Studies.

Davis, A., Murphy, J., Owens, D., Khazanchi, D., & Zigurs, I. (2009). Avatars, people, and virtual worlds: Foundations for research in metaverses. *Journal of the Association for Information Systems*, *10*(2), 91-117.

Davis, J. W. (2006). Our Achilles' Heel: Language Skills. *Military Review*, *86*(2), 110-111.

DeChoudhury, M., Mason, Winter A., Hofman, J.M., & Watts, D.J. (2010). Inferring Relevant Social Networks from Interpersonal Communication. *Proceedings of the Nineteenth International Conference on World Wide Web*, 301-310, Raleigh, NC. doi: http://dx.doi.org/10.1145/1772690.1772722

Ding, S., Cong, G., Lin, C.Y., & Zhu, X. (2008). Using conditional random field to extract contexts and answers of questions from online forums. *Proceedings of 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 710–718, Columbus, OH.

Dombroski, M., Fischbeck, P., & Carley, K. (2003). Estimating the shape of covert networks. *Proceedings of the 8th International Command and Control Research and Technology Symposium*, National Defense War College, Washington, DC.

Drain, B. (2010). The largest battle ever held in EVE Online is going on right now. Retrieved August 14, 2016, from https://www.engadget.com/2010/10/30/the-largest-battle-ever-held-in-eve-online-is-going-on-right-now

Fisher, D.R., Waggle, J., & Leifeld, P. (2013): Where does Political Polarization Come From? Locating Polarization Within the U.S. Climate Change Debate. *American Behavioral Scientist*, *57*(1): 70-92.

Freeman, L.C. (1979). Centrality in social networks: conceptual clarification. *Social Networks*. *1*(3), 215-239.

Gordon, D. & Lakoff, G. (1971). Conversational postulates. *Proceedings of the Seventh Regional Meeting of the Chicago Linguistic Society*. Chicago, IL: Chicago Linguistic Society, 63-84.

Goyal, S., Van Der Leij, M. J., & Moraga-Gonzalez, J. L. (2006). Economics: An emerging small world. *Journal of Political Economy*, *114*, 403-412.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, *78*, 1360-1380.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (eds.), *Syntax and Semantics Volume 3: Speech Acts*, New York: Academic Press, 41-58.

Herring, S. (2004). Computer-mediated discourse analysis: An approach to researching online communities. In S.A. Barab, R. Kling, J.H. Gray (eds.), *Designing for Virtual Communities in the Service of Learning*, New York, NY: Cambridge University Press, 338–376.

Hong, L., & Davison, B. D. (2009). A classification-based approach to question answering in discussion boards. *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 171-178, New York, NY, Association for Computing Machinery.

Hummon, N. P. (2000). Utility and dynamic social networks. *Social Networks*, *22*, 221-49.

Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (eds.), *Sociolinguistics: Selected Readings*. Harmondsworth: Penguin, 269-293.

Johnstone, B. (2000). *Qualitative Methods in Sociolinguistics*. New York, NY: Oxford University Press.

Johnstone, B. (2008). *Discourse analysis*. Malden, MA: Blackwell.

Kalou, Z., & Sadler-Smith, E. (2015). Using Ethnography of Communication in Organizational Research. *Organizational Research Methods*, *18*(4), 629-655.

Kwong, H., & Yorke-Smith, N. (2012). Detection of imperative and declarative question–answer pairs in email conversations. *AI Communications*, *25*(4), 271-283.

Labov, W. & Fanshel, D. (1977). *Therapeutic Discourse*. New York, NY: Academic Press.

Leifeld, Philip & Haunss, S. (2012). Political Discourse Networks and the Conflict over Software Patents in Europe. *European Journal of Political Research, 51*(3): 382-409.

Lofland, J., Snow, D., Anderson, L., & Lofland, L. (2006). *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis (4th edition)*, Belmont, CA: Wadsworth/Thomson Learning.

Magouirk, J., Atran, S., & Sageman, M. (2008). Connecting terrorist networks. *Studies in Conflict & Terrorism*, *31*(1), 1-16.

Malm, A., Bichler, G., & Nash, R. (2011). Co-offending between criminal enterprise groups. *Global Crime*, *12*(2), 112-128.

Malm, A., & Bichler, G. (2011). Networks of collaborating criminals: Assessing the structural vulnerability of drug markets. *Journal of Research in Crime and Delinquency*, *48*(2), 271-297.

Messinger, P. R., Stroulia, E., Lyons, K., Bone, M., Niu, R. H., Smirnov, K., & Perelgut, S. (2009). Virtual worlds—past, present, and future: New directions in social computing. *Decision Support Systems*, *47*(3), 204-228.

Milgram, S. (1967). The small world problem. *Psychology Today*, *2*(1), 60-67.

Milroy, L. (1987). *Language and Social Networks. 2nd Edition*. Baltimore, MD: University Park Press.

Moser, C. Groenewegen, P. & Huysman, M. (2013). Extending Social Network Analysis with Discourse Analysis: Combining Relational with Interpretive Data. In, T. Özyer, J. Rokne, G. Wagner, A.H.P. Reuser (Eds.), *The Influence of Technology on Social Network Analysis and Mining, Lecture Notes in Social Networks, 6*, 547-561.

Newman, M.E.J., Barabasi, A.L., & Watts, D.J. (2006). *The Structure and Dynamics of Networks.* Princeton, NJ: Princeton University Press.

Paltridge, B. (1998). Get your terms in order. In P. Master & D. Brinton (eds.), *New Ways in English for Specific Purposes*. Alexandra, VA: TESOL, 263-266.

Paltridge, B., & Burton, J. (2000). *Making sense of discourse analysis*. Gold Coast.

Paltridge, B. (2012). *Discourse analysis: An introduction.* New York, NY: Bloomsbury Publishing.

Pfeffer, J., & Carley, K. M. (2012). Rapid modeling and analyzing networks extracted from pre-structured news articles. *Computational and Mathematical Organization Theory*, *18*(3), 280-299.

Potter, J. (1996). Discourse analysis and constructionist approaches: Theoretical background. In J.T.E. Richardson (ed.), *Handbook of Qualitative Research Methods for Psychology and the Social Sciences*, 125-140, Leicester: British Psychological Society.

Putnam, R. D. (1993). *Making Democracy Work: Civic Traditions in Modern Italy.* Princeton, NJ: Princeton University Press.

Ressler, S. (2006). Social network analysis as an approach to combat terrorism: past, present, and future research. *Homeland Security Affairs*, *2*(2), 1-10.

Robins, G., Pattison, P. & Woolcock, J. (2005). Small and other worlds: Global network structures from local processes. *American Journal of Sociology*, *110*, 894-936.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*, 696-735.

Sageman, M. (2011). *Leaderless jihad: Terror Networks in the Twenty-First Century*. Philadelphia, PA: University of Pennsylvania Press.

Shapiro, J. N. (2013). *The Terrorist's Dilemma: Managing violent covert organizations*. Princeton, NJ: Princeton University Press.

Shirky, C. (2008). Here Comes Everybody: *The Power of Organizing without Organizations*. New York, NY: The Penguin Group.

Sivan, Y. (2008). 3D3C real virtual worlds defined: The immense potential of merging 3D, community, creation, and commerce. *Journal of Virtual Worlds Research*, *1*(1), 1-32. doi: http://dx.doi.org/10.4101/jvwr.v1i1.278

Smith, M., Milic-Frayling, N., Shneiderman, B., Mendes Rodrigues, E., Leskovec, J., & Dunne, C. (2010). *NodeXL: a free and open network overview, discovery and exploration add-in for Excel 2007/2010*, Retrieved from http://nodexl.codeplex.com

Stephenson, K. A., & Zelen, M. (1989). Rethinking centrality: Methods and examples. *Social Networks*, *11*, 1–37.

Sun, L., Liu, B., Wang, B., Zhang, D., & Wang, X. (2010). A study of features on Primary Question detection in Chinese online forums. *Proceedings of Seventh International Conference on Fuzzy Systems and Knowledge Discovery*, Yantai Shandong: IEEE, *5*, 2422-2427. doi: http://dx.doi.org/10.1109/FSKD.2010.5569298

Swales, J. M. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge, MA: Cambridge University Press.

Tsvetovat, M., & Carley, K.M. (2007). On effectiveness of wiretap programs in mapping social networks. *Computational and Mathematical Organization Theory*, *13*(1), 63-87.

Van Dijk, T. A. (1993). Principles of critical discourse analysis. *Discourse & Society*, *4*(2), 249-283.

Xu, J., & Chen, H. (2008). The topology of dark networks. *Communications of the ACM*, *51*(10), 58-65.