In this paper, we explore the flow of information regarding strategic Middle Eastern countries in the WikiLeaks ‘diplomatic cables’ by applying data-mining techniques to construct directed networks. The results show that between 2002 and 2009, US diplomatic communication related to these countries increased although with notable variation in flow patterns. We discuss the value of a visual display of diplomatic communication patterns in understanding the decentralized nature of information gathering on regional foreign policy issues.

Introduction

Diplomatic missions perform an essential function in their contributions to the process of developing policy on global and regional issues. A century ago, a state had a limited number of diplomatic missions, but in the contemporary world, a few geographically dispersed missions would not be sufficient to carry out these functions. Scholars, however, have had limited access to information on the mechanisms and frequency of diplomatic communications. Given this lack of access, research into the flow of diplomatic information between a country’s centre and its diplomatic missions has not previously been able to visualize the actors and structure of communication. The construction of the network of a state’s diplomatic correspondence on a specified region would open a new path for further analysis of foundations of foreign policy development.

This paper builds on a long line of research espousing the value of network analysis in international relations. The release of the WikiLeaks documents, known as ‘diplomatic cables’, allowed us to explore the information flows between the US State Department and US diplomatic missions. After selecting the Middle East region as a case study, we asked whether it is possible to visualize the diplomatic focus on these Middle Eastern countries between 2002 and 2009 by mining the information in the documents. To answer this question, we constructed directed networks which display the countries of the diplomatic missions sending the cables as sources and the Middle Eastern countries about which information was sent as targets.
The resulting directed networks illustrate that an increasing number of US diplomatic missions sent cables mentioning these Middle Eastern countries during the period under review, but the patterns of communication varied. First, the diplomacy networks reveal an increasing global interest in the US presence in the region. Second, the diplomatic missions competed in contributing information, decentralizing the distribution of information. While it can be assumed that diplomatic missions would be likely to send information on neighbouring countries, finding that far distant countries contributed to the Middle East communication network reveals the dynamism of US diplomatic communications. Third, the shifting focus of communications shows that the weight of the Middle East countries has shifted within the US diplomacy network.

Our work contributes to the literature by developing a new source of data and employing network analysis to visualize a complex set of interactions between the US State Department, foreign embassies and consulates. The novelty of our analysis comes from the visualization of diplomatic communications which previously had been closed to analysis. In essence, we propose a new and surprisingly simple technique to grasp the scale of information disclosed within the ‘WikiLeaks diplomatic cables’.

**Literature Review**

The whistle-blower organization WikiLeaks emerged in the headlines of the international press in July 2010 with the release of classified US military documents related to the Afghan war, a set of information now known as the ‘Afghan War Logs’. Soon after the ‘Afghan War Logs’, the organization released approximately 400,000 documents on the Iraq war and 250,000 diplomatic cables, which had been sent between the US State Department and the US foreign missions. By the end of 2010, the number of documents released by the website had reached nearly 730,000, a historically unprecedented amount of classified information. The release of the cables presented the scholarly communities in international relations, intelligence gathering and journalism studies with a rich source of information related to the contemporary history of international relations.

Since the disclosure of the documents, scholars have examined them with varying concerns, including the authenticity of the cables. Some argued that disclosure of the documents was irresponsible for security reasons, others found the sheer volume of the cables an anomaly, or in contrast, dismissed them as only a small part of US foreign policy communications and thus being given a disproportionate significance in public discourse. Several studies focused on the significance of the cables for US diplomacy, the future of intelligence gathering and whether the leaked documents could be considered as ethical journalism. The majority of the studies, however, tended to focus on the implications of the disclosures for transparency in international relations and in US domestic politics. An immediate reaction, for instance, concerned the potential role of the documents in promoting transparency against corruption and the role of citizens as watchdogs.

Despite these different opinions as to the political implications of the WikiLeaks documents, there has been little systematic structural analysis of the cables, by which we mean a method that goes beyond the context analysis of a subset of cables.
Two studies using the Afghan War Logs as their primary source of data are the exceptions. The first study codes the events and casualties in the disclosed documents in comparison with the military and news-based data.\textsuperscript{12} In the second study, mortality rates are modelled by mining the text in the Afghan War Logs.\textsuperscript{13} We take a similar data-mining approach to the ‘WikiLeaks diplomatic cables’ as a quantitative data source, and propose that it is possible to display the information flows scattered across the documents and tease out the systematic connections among them.

**Structural Characteristics of the WikiLeaks ‘Diplomatic Cables’**

Most (if not all) of the WikiLeaks ‘diplomatic cables’ have several common characteristics allowing researchers to compare and classify them. In the majority of cases, the document structure is uniform and can be divided into three parts, each of which can be examined separately.

The first part of each document identifies the diplomatic mission sending the cable. All of the cables begin with a header, which contains the cable ID (85MOSCOW8814 or 08BAMAKO799). Each WikiLeaks diplomatic cable has a unique identification number, which is displayed as the first item on the first line of each document (09ISTANBUL290 or 04BEIRUT408). A time stamp (1/22/2004 16:29) follows the identification number. The documents clearly state the origin of the diplomatic cable, such as ‘Embassy Beirut’, and the addresses of the recipient duty stations follow the address of the sender. The location of the recipient is also noted with a dispatch code to track the report, if necessary. The priority of the distribution locations is also stated, with the possibility of more than one priority address. We describe the detailed structure of the cables in the Appendix.

The most important attribute of the documents is the degree of secrecy. The highest level of security classification of the material is ‘Top Secret’. Top Secret cables are not among those released on the Internet, despite initial reports in the media suggesting otherwise. Some of the disclosed diplomatic cables are categorized as ‘Secret’, a classification which signals that such material would cause ‘grave damage’ to US national security if it were publicly available. The other category, which is the most frequent, is ‘Confidential’, which signals that the subject matter would cause ‘damage’ or be ‘prejudicial’ to US national security. Both categories are further labelled on the basis of access of Foreign Nationals. For instance, Confidential//NOFORN is the same as the ‘Confidential’ category, with NOFORN meaning the dissemination is restricted to US nationals. Similarly, Secret//NOFORN signals the document is secret and restricted to US nationals.

The category ‘Unclassified’ is technically not a classification level, but is used for government documents which do not carry one of the classifications listed above. Such documents can sometimes be viewed by people without a security clearance. An additional category, Unclassified//FOR, allows dissemination of the document to foreign nationals. Finally, ‘Official Use Only’ signals sensitive information that is restricted.

How each document is to be classified, safeguarded and declassified as national security information is determined by Executive Order of the US Presidency.
The diplomatic cables begin with the initials of the Executive Order followed by a number such as 12958. Additionally, each document has a different date for declassification, depending on its content. In some documents, the declassification dates are not available, which is noted by the abbreviation N/A. The declassification date is followed by tags.

The tags in diplomatic cables vary by country, organization, policy program and subject. All country tags follow the US Federal Information Processing Standards. The organization tags that appear in this section represent international or domestic organizations. Most of the acronyms are formed by the initials of organizations, such as AEC for Atomic Energy Commission, DEA for Drug Enforcement Administration and JUS for Department of Justice. In addition, the cables carry subject tags, such as Employees Abroad (AODE), Energy and Power (ENRG) or Trade and Technology Controls (ETTC).

The third and last part of each cable is the body of the message. This part consists of titles, excerpts, interviews and subtitles.

**Construction of Network Data by Mining the Diplomatic Cables**

Algorithms to mine massive volumes of text to construct numeric data for visualization have been refined over the years. We apply this methodology as follows: first, the diplomatic cables were downloaded from the WikiLeaks website (http://wikileaks.org/cablegate.html). The archive contains a single compressed text file. After extracting and decomposing the single file, we derived the original 251,287 documents, of which 15,652 were secret; 101,748 confidential; and 133,887 unclassified, originating from 274 different diplomatic missions.

For our case study, we narrowed our search to a single geographical region, the Middle Eastern countries: Bahrain, Egypt, Iran, Iraq, Israel, Jordan, Lebanon, Kuwait, Qatar, Saudi Arabia, Syria, Turkey and Yemen. Scanning the documents one at a time, the data-mining algorithm extracted the following information from the first part of each diplomatic cable: (i) the cable unique identification number; (ii) the date and time the cable was sent; (iii) the code which includes the year and embassy; (iv) the country of the reporting duty station; and (v) the destination embassy and consulates of the cable. To facilitate comparison, the embassy and consulate names are mapped to their associated countries. From the second part of each cable, the algorithm mined the type of cable, such as Unclassified or Confidential.

The data algorithm then searched the contents of the cable to count the number of times the name of any of the selected Middle Eastern countries appeared in the text. Our decision to count the number of appearances of the country names stems from the simple hypothesis that the greater the number of references to a particular country, the higher the importance of that country in that particular diplomatic cable. Therefore, while our approach does not involve context analysis, the data-mining algorithm calculated the number of country mentions as the weight, that is, importance.

Finally, for each year between 2002 and 2009 and for each country pair (Country sending the document, Country discussed in the document), the algorithm summed up the number of references to the discussed country in the cables and constructed
the source data fed into the network analysis as follows: in the year \( x \), the cables sent from \( \text{Country1} \) contains \( y \) times mentions of \( \text{Country2} \). \( \text{Country1} \) can be any country, but \( \text{Country2} \) must be one of the selected Middle Eastern countries.

Network analysis is a method that is used to identify and visualize interactions. The interactions are schematically represented by nodes, which may represent individuals, organizations or entities, and links, which represent the direction and frequency of contacts among the nodes.\(^{15}\) The graphic representation of interactions provided by network analysis allows researchers to visually examine the topology of the network, facilitating the identification of relationships among the actors, which may not be as apparent in numerical representation or written description. By plotting the vectors of contact, a network analysis creates a structured representation of a set of relationships. Network analysis has been applied to the study of many different types of relational interaction from the dynamics of terrorist networks to the spread of disease in epidemiology to the migration dynamics of human populations.\(^{16}\)

Given that the diplomatic cables are the means of transmitting information, we chose to represent the relational structure of the information flow with directed networks. Each node represents a country: the source node is the country where the cable originated and the target node is a Middle Eastern country, which is mentioned in the document. As stated above, the target countries in our analysis are Bahrain, Egypt, Iran, Iraq, Israel, Jordan, Lebanon, Kuwait, Qatar, Saudi Arabia, Syria, Turkey and Yemen. The source nodes are all countries hosting a US foreign mission, including these Middle Eastern countries, as self-reporting is to be expected. The links reflect the direction of the information flow from the reporting country to the Middle Eastern country reported on. We began our analyses with the cables sent in 2002 in the immediate aftermath of the September 11 attacks and ended with the cables sent in the year 2009.

The networks can be represented by different measures, such as degree of each node and clustering coefficient of the network. The most fundamental characteristic of a node is its degree (or connectivity), \( k \), which is the number of links to that one node. In undirected networks, the degree shows the number of links connected to a particular node and the average degree of a network is calculated by dividing the total number of interactions by the total number of nodes, \( \langle k \rangle \). In directed networks, however, the value of the degree can differ based on the direction of the interaction (incoming or outgoing). In our case, for each Middle Eastern country, the in-degree denotes the number of countries that send cables about that country, while out-degree denotes the number of Middle Eastern countries mentioned by the reporting US mission in that Middle Eastern country. For example, in 2002, information about Bahrain was sent by the US foreign missions in 47 different countries (i.e. in-degree). Hence, in-degree indicates ‘popularity’. The out-degree in our model shows low volatility as our sample comprises 13 Middle East countries. For Bahrain, in 2002 the out-degree is 13, indicating that the US diplomatic mission in Bahrain reported on all of the countries included in the sample. In our model, the in-degrees, that is, the number of countries with US diplomatic missions reporting on the target country indicates the degree of focus on that country. As some of the documents would contain self-reporting, we allowed for self-referential loops in the networks that we constructed.
The links in the directed networks display the total number of references to each target Middle Eastern country in the cables sent by each source country and determine the link weight, a value indicating the relative importance of the reported country. The link weight is indicated in the network visualizations by the thickness and style of the link: the darkness and thickness of the arrow increase with the frequency of the appearance of a particular Middle East country in cables sent from the reporting country. A thick solid line represents 10,000 or more references, while a thin dotted line represents as few as 10. A high frequency of references to a specific country is thus clearly distinguished from a low frequency in the network representations.

The clustering coefficient of a node indicates the cohesiveness of neighbouring nodes. In a directed network, a node with \( k \) neighbours can have at most \( k(k - 1) \) links (this value would indicate that every neighbour of this node is connected to every other neighbour). The clustering coefficient is a value ranging between zero and one, where one indicates all neighbours of a particular node are connected to each other.\(^{17}\) While the clustering coefficient is a property of each node, the average clustering coefficient of a network is the average of all clustering coefficients, denoted by \( \langle C \rangle \). The highest average clustering coefficient of a network, 1, indicates that all nodes in the network are connected to each other.

**Discussion of the Flow of Information as Documented in the Cables**

A set of directed networks linking reporting duty stations to the Middle Eastern countries for each year between 2002 and 2009 are depicted in Figure 1. The target Middle Eastern countries are located at the centre in a circle to simplify visualization. In 2002, a relatively small number of US diplomatic missions sent diplomatic cables reporting on the Middle Eastern countries. Some of these reporting countries, as expected, are located in the Middle East region, while others are as remote as Brazil and Sri Lanka. A visual scan over the years in our sample shows an expanding volume.

![Figure 1](image-url)
of cables on these countries by US diplomatic missions. Given the political, social and economic developments of the last decade in the Middle East, it is not surprising that an increasing number of US diplomatic missions would send an increasing number of cables on these countries.

In addition to the visual representation, two statistics that summarize the changes in network topology over the years are also displayed in Figure 1. The overall degree of the network increased from four to eight, indicating that on average in 2002 all duty stations reported on about four Middle Eastern countries. By 2009, the average number of Middle Eastern countries reported on had risen to eight. The clustering coefficient, which had risen from 0.4 to 0.9, also indicates an increasing information flow among the duty stations. The clustering coefficient tells us that by 2009 the network was almost completely connected and all of the US foreign missions reported on the selected Middle Eastern countries.

We then assembled Figure 2, which displays the summarized in- and out-degrees of the directed networks. In all of the countries included in the sample, in-degree, that is, the number of US diplomatic missions reporting on the Middle Eastern countries, steadily increased from 2002 to 2009, with a significant jump in 2006. The only exception is Lebanon, where the in-degree decreased following a peak in 2006 when Rafiq Hariri, a former Prime Minister, was assassinated. The in-degrees indicate that Israel, Egypt, Iraq and Iran were reported on most frequently. We thus consider these countries to be the first-tier group. The second tier of countries includes Jordan, Kuwait, Lebanon, Saudi Arabia, Syria and Turkey. The Middle East countries reported on by the fewest foreign missions are the Gulf countries, Qatar and Bahrain, in addition to Yemen.

The out-degree distributions cannot exceed the sample size, a total of 13 Middle East countries. One observation is notable: until 2005, the out-degree of Iran was zero, due to the lack of a US foreign mission in Iran. However, by 2006, out-reporting from Iran had increased. By checking the relevant cables, we found that the USA operates a duty station (Iran Regional Presence Office) located in Dubai in the United Arab Emirates (UAE) for Iran, which is the source of the out-degree as a result for Iran. As expected, the out-degrees for Jordan, Kuwait and Turkey between 2002 and 2009 was 13, indicating steady generation of reports on Middle Eastern countries, including themselves.

While the in-degree and out-degree distributions for 2002–2009 show clear trends in the ‘popularity’ of the target country and the ‘activity’ of the US diplomatic mission in that Middle Eastern country, counting only the number of reporting countries would result in a significant loss of information. For example, in 2007, the US diplomatic missions in 52 countries reported on Bahrain, by 2009, the number had increased to 54 countries (in-degree). However, in the same batch of cables, Bahrain was mentioned 2586 times in 2007 and 2213 times in 2009 of which 2234 and 1548 were self-mentions, respectively.

We addressed this difference between the number of US foreign missions reporting and the number of country references in the texts by counting the name of a particular country in the diplomatic cables. We then divided the total number of references (NoR) to a country in the diplomatic cables into self-reporting (Self) and reporting by other countries (Net).
Figure 3 displays the trends in self-reports, other reports and total reports of a target country in the cables. Again, we see three distinct groups, in the total number of times each country is named in the cables. The countries in the region most frequently reported on are Iran, Iraq and Israel, however, the patterns vary. In 2002, Iran was named in approximately zero cables. Between 2005 and 2006, a significant increase is observed in the number of references to Iran in cables sent by other
countries. Self-reporting on Iran is constantly low with an increase in 2007, likely a result of opening the Dubai station in UAE. Note that the number of reports on Iran in the cables is the highest among all Middle Eastern countries (over 40,000/year after 2005). The volume of cables on Iraq and Israel closely followed Iran. The number of self-reports originating in Israel, closely tracks the number of times it is reported on in cables sent by US foreign missions in other countries. The second

Figure 3 Directed Networks of WikiLeaks Diplomatic Cables, Self-, Other and Net Mentions
group based on number of references includes Egypt, Kuwait, Lebanon, Syria and
Turkey. The third group of least referenced countries compared to the other countries
of the region comprises of Bahrain, Qatar, Saudi Arabia and Yemen.

It is also possible to calculate the host and foreign activity rate: in which countries
did the US foreign mission report on the host country more frequently than foreign
missions located in other countries? Figure 3 divides the countries into two groups:
those in which the US foreign mission reported on the host country more frequently
than foreign missions located in other countries include Bahrain, Jordan, Kuwait,
Qatar, Turkey and Yemen. The similar self- and total number of references indicates
that the majority of the cables generated from these countries were predominantly
self-reporting. In contrast, Syria, Saudi Arabia, Lebanon, Iraq and, understandably,
Iran are more frequently reported on in cables sent by diplomatic missions in other
countries. These trends highlight the geographical dispersion of US foreign missions
which have the capacity to help in shaping the US foreign policy.

Conclusion

While methods for the visualization of data are familiar since the emergence of
cartography, recent advances in computer technology now allow the quantification
and display of textual content at detailed frequency levels. As a result, we are now able
to better represent information flows by consolidating characteristics that would
otherwise be scattered among a myriad of documents. The WikiLeaks documents
provide a unique opportunity to understand the patterns of country-level
information flow in modern US diplomacy, one of the largest diplomatic networks
in existence.

In applying network analysis to international relations, this paper extends the work
of Hafner-Burton et al. (2009) which argues that network analysis helps to identify
relational structures among actors. Here we constructed a network of the information
flow revealed in the WikiLeaks diplomatic cables between US foreign missions using
the Middle East region as a case study.

Our findings suggest that visualization can help us to better understand the flow of
information in the WikiLeaks documents, which have thus far not been analysed
systematically. We demonstrate that while the US diplomatic missions located in the
Middle East predominantly reported on neighbouring and non-neighbouring
countries, diplomatic missions located in countries outside of the region also
followed developments in the region. We did not find that all of the Middle Eastern
countries selected for this study self-reported or reported on the other Middle
Eastern countries in similar patterns. Our work does not attempt to draw any
conclusions in regard to US foreign policy, but presents a novel way to build a
foundation for new questions in the scholarly literature on international relations.

Admittedly, our analysis raises more questions than it answers: is it possible that a
similar trend could be observed in other regions such as Latin America or South Asia?
Would other network analysis tools yield similar results? Why would US missions
outside of the Middle East region compete in collecting information on the region?
Future research in this area could compare our results with another region, such as
South East Asia. Other network methods could reveal more about the structure of networks built on the information flow shaped by similar documents. Another interesting research area would be to investigate competition among the nodes, that is, diplomatic missions competing within the diplomacy network. For example, a future study utilizing network analysis could compare the patterns observed in in-degree values with the timeline of regional events. Such an analysis would provide new and important insights for scholars of international relations.

Notes


A. Salih Bicakci (corresponding author) is currently a full-time associate professor at the Department of International Relations of Kadir Has University, Turkey. He received his PhD degree from the Department of Middle Eastern and African History at Tel Aviv University, Israel. He has had articles published on the security and identity of the Middle East. His recent research focuses on cyber security and threats.

Address for correspondence: Kadir Has Üniversitesi, Kadir Has Caddesi, Cibali, 34083 Istanbul, Turkey. E-mail: asbicakci@khas.edu.tr and asbicakci@gmail.com

Deniz Rende is currently working as a postdoctoral researcher at Rensselaer Polytechnic Institute, USA. She received her PhD degree from the Department of Chemical Engineering at Bogazici University, Turkey. Her current research focuses on systems biology approaches to human diseases and modular structures in biological networks.

Address for correspondence: Rensselaer Polytechnic Institute, Materials Research Center Room 311, RPI 110 8th Street, MRC 140, Troy, NY 12180-3590, USA. E-mail: rended3@rpi.edu
Sevinc Rende is currently a full-time assistant professor at the Department of Economics, Isik University, Turkey. She received her PhD degree from the Department of Economics at the University of Massachusetts Amherst, USA. Her academic work maps conflict and social networks in post-conflict and fragile states.

Address for correspondence: Isik University, Kumbaba Mevkii Sile, 34398 Istanbul, Turkey. E-mail: rende@isikun.edu.tr

Olcay Taner Yildiz is currently a full-time associate professor at the Department of Computer Science and Engineering, Isik University, Turkey. He received BS, MS and PhD degrees in computer science from Bogazici University, Istanbul, in 1997, 2000 and 2005, respectively. He worked extensively on machine learning, specifically model selection and decision trees. His current research interests include software engineering, natural language processing and bioinformatics.

Address for correspondence: Isik University, Kumbaba Mevkii Sile, 34398 Istanbul, Turkey. E-mail: olcaytaner@isikun.edu.tr
Appendix

The first part of a WikiLeaks document includes the release date (such as 9/17/2008 16:17) of the document. The classification categories follow the time stamp on the document. In the recipient category, there are abbreviations, such as RUEHC. SECSTAT WASHDC PRIORITY 9647, RUEHC INFO RUCNDT/USMISSION USUN NEW YORK PRIORITY 0247. The State Department always used the RUEHC code. Other rules on revising and drafting new materials are elaborated in the Foreign Affairs Manual Standards file, see http://www.state.gov/documents/organization/89605.pdf

Top Secret classification is the highest level of secrecy on a national level. As expected, the protection of the document has vital importance. According to Executive Order 11652, ‘unauthorized disclosure [of such material] could reasonably be expected to cause exceptionally grave damage to national security’. It would also compromise the defence strategies of the USA. Therefore, the Top Secret classified documents were sorted out of the WikiLeaks collection. In the early days of the WikiLeaks release, the Internet shook with the news that the Top Secret material would harm US national security. Later, it was clarified that there was no document in the Top Secret classification category.

In earlier regulations, the US State Department was explaining declassification and downgrading the procedure as ‘Top Secret’ shall become automatically downgraded to ‘Secret’ at the end of the second full calendar year following the year in which it originated. However, US President Obama amended the classification with the new Executive Order 13526 and authorized the classification authority to ‘establish a specific date or event for declassification based on the duration of the national security sensitivity of the information’. If the classification authority does not set a date, then the document would be declassified in a minimum of 10 years and a maximum of 25 years. In the Executive Order, it is noted, ‘no information may remain classified indefinitely’. The material published on the net, as a result, does not contain documents under the ‘Top Secret’ classification (see Executive Order 11652, http://www.fas.org/irp/offdocs/eo/oe-11652.htm, Executive Order 11652, http://www.fas.org/irp/offdocs/eo/oe-11652.htm, Executive Order 13526, http://edocket.access.gpo.gov/2010/pdf/E9-31418.pdf, Executive Order 13526, http://edocket.access.gpo.gov/2010/pdf/E9-31418.pdf).

Following the release of the WikiLeaks diplomatic cables, the management of security information was regulated with Executive Order 13587. The declassification date (such as 05/07/13) is also followed by the classification Executive Order number. See Executive Order 13587, http://www.whitehouse.gov/the-press-office/2011/10/07/executive-order-structural-reforms-improve-security-classified-networks-

The country tags used in construction of the networks include regional tags based on geographic locations. For instance, the following tags, AE for United Arab Emirates, GZ for Gaza Strip, IS for Israel, denote the countries. For the regional tags, XA Africa, XC South East Asia, XF Middle East, XG Europe, XH Eastern Europe are used to define the region (source: Federal Information Processing Standards Publication 10-4, 1995 April, Announcing the Standard for Countries, Dependencies,

From time to time, the subject tags are close to the program tags, which are abbreviations usually starting with K, such as KACT for Strategic Arms Control (ACS) Treaties, KCIP for Critical Infrastructure Protection. In some documents there are exceptions such as KWBG for West Bank and Gaza that shows National Authority of Palestine. In addition to these tags, there are some unidentified tags, such as MTRE, OTR and PBIO.