US-Russia Cooperation: Understanding Cybersecurity Contextualizations

Ashe Magalhaes[1], Elza Ganeeva[2], and Susan Wen[3]

Abstract— **This paper aims to mitigate the consequences of false attribution and ensuing crisis escalation between the U.S. and Russia by conducting a a) policy analysis that explores the current state of cybersecurity cooperation across three levels: unilateral, multilateral, and multi-stakeholder and by b) examining a tool that identifies the most discordant cybersecurity-related words used between English and Russian speakers. The contribution of the paper is a novel approach to US-Russia cooperation in cyberspace that establishes a foundation for understanding the differences in how English and Russian speakers contextualize cybersecurity.**

I. INTRODUCTION

According to a survey of the Pew Research Center released in October 2014, about two-thirds of technology experts expect a major cyber-attack in the world by 2025, which is likely to lead to significant loss of life or property (Rainie 2014). The threat of cyber operations places an asymmetric burden on a passive defense strategy, as the attacker need only exploit the system once to cause significant damage. As a result, the international debate over cybersecurity necessarily includes a consideration of attack options for defensive purposes (Owens et al. 2009, 14). This offensive strategy, along with the difficulty of establishing cyberattack attribution, presents the problem of false attribution and ensuing crisis escalation.

The United States (US) and Russia are particularly vulnerable to conflict escalation resulting from the false attribution of cyber attacks due to weak communication channels. This, coupled with the military capabilities of both countries, motivates cooperation in cyberspace. We hypothesize that the problem of false attribution and ensuing crisis escalation between the US and Russia can be mitigated by stronger communication channels. Strong communication channels, in turn, can be developed by qualitative and quantitative approaches to understanding how the US and Russia each contextualize cybersecurity.

The paper is divided into two sections: **Policy Analysis: The Current State of Cybersecurity Cooperation** and **Empirical Analysis: An Identification of Discordant Cybersecurity Terms**. The policy analysis provides an overview of the existing cybersecurity landscape and its limitations; the empirical analysis details a novel and technical approach to practically enhancing communication. Both approaches are necessary to develop a new strategic framework that prevents a breakdown of communication between the American and Russian governments. By bringing to light failure points in the way American and Russian government officials speak to each other about cybersecurity-related events, we offer both governments a better chance at cooperation.

II. POLICY ANALYSIS: THE CURRENT STATE OF CYBERSECURITY COOPERATION

This section explores the policy challenges of cybersecurity, cybersecurity policy organization on the global stage, and the current state of American and Russia cooperation in cyberspace.

A. Cybersecurity Policy Challenges

1) The Shared Threat of Compromised Industrial Control Systems (ICS) : In both the US and Russia, the vulnerability of ICS threatens large scale disruption in core infrastructure such as dams, energy systems, transportation systems.

[1]Ashe Magalhaes, Stanford University Department of Computer Science

[2]Elza Ganeeva, Higher School of Economics, Governance in Science Technology and Innovation

[2]Susan Wen, Rice University Department of Computer Science

Cyberattacks on these systems cause physical damage. Most ICS components, such as Programming Control Logic (PLC) and network devices, were built on the assumption that the systems operate in isolated environments (Andreeva, et al. 2016, 3). As more ICS become connected to the Internet, adversaries are able to exploit this design flaw. For example, buffer overflow attack involves an adversary taking advantage of the network services' root access to the operation system, thus granting the remote adversary control of the machine.

The threat of compromised ICS is worsened by the fact that these systems are usually embedded in physical machines. These machines have a lifespan of ten to fifteen years, as opposed to the three to five year lifespan of an information technology system (Stouffer et al. 2015, 3). As a result, even when ICS are detected, they are difficult to fix. According to IBM Managed Security Services (MSS) data, attacks targeting ICS increased over 110 percent in 2016 over last years numbers (McMillen, 2016).

2) Difficulty in Establishing Deterrence in Cyberspace : The success of global nuclear nonproliferation has inspired scholars to explore the analogy of cyberattacks to nuclear weapons. However, the fundamental differences between the two domains highlight the difficulty in establishing deterrence in cyberspace. Deterrence includes two components: deterrence by punishment (the threat of retaliation) and deterrence by denial (the ability to prevent benefit) to disincentivize hostile actions (Jasper 2015, 61). The ability of nuclear weapons to change the balance of power on the battlefield and ensure decisive victory through assured destruction granted nuclear power its credibility (Cirenza, 2015, 9). The perceived potential for destruction from nuclear weapons quenched the incentive for offenses from third party adversaries (Levy, 1984). Even without the threat of mutually assured destruction, nuclear materials are not easy to obtain.

In contrast, cyber weapons can be acquired through research and development by either state or non-state actors. The success of cyberattacks relies on secrecy rather than the ability to destroy because once victims realize they are vulnerable, they can build patches to undermine future cyber offenses. For deterrence by punishment to be successful, states must be able to carry out self-defense. Yet, there is no established consensus on necessary and proportionate response in cyberspace (Jasper 2015, 66). Unlike in nuclear warfare, escalation in cyberspace does not provide for mutually assured destruction. Difficulty in cyberattack attribution complicates the justification of self-defense. **The nuclear analogy does not hold because cyberspace does not deter hostile actions through deterrence by punishment or deterrence by denial.**

B. Current Global Efforts in Establishing Confidence Building Measures

On the global stage, cybersecurity solutions are organized into unilateral, multilateral, and multi-stakeholder approaches. On the unilateral level, countries enforce a legal framework on cyber activities. On the multilateral level, states enact treaties and agreements among themselves, identify acceptable and illegal actions, and propose norms and cooperations. On the multi-stakeholder level, states engage private sectors and academics. Below is an outline of existing confidence building measures implemented on the multilateral and multi-stakeholder levels.

1) Multilateral Approach: Most influential multilateral efforts in cybersecurity cooperation come from a series of reports from the United Nations Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security (UN GGEs). UN GGE discusses international norms on cybersecurity in detail and proposes confidence-building measures. The UN GGE recommends that states create a directory for sharing points of contact in addressing cybersecurity emergencies, encourages transparency in sharing points of views on information technology, and suggests that states strengthen cooperative mechanisms by encouraging research institution exchanges (UN GGE 2015).

Another example of international multilateral efforts is the US and Russia's bilateral agreement of 2013, which pioneered bilateral efforts to establish shared trust between states. The agreement details that confidence building measures for the two countries become a major focus and encourages

stronger collaboration between the national Computer Emergency Response Teams (US-CERT and RU-CERT) through regular information sharing. The agreement includes the establishment of a communication hotline between the White House and the Kremlin in case of a cybersecurity crisis. It also expands the responsibility of Nuclear Risk Reduction Centers in both US and Russia to incorporate cybersecurity incidents.

Multilateral efforts include the following:

- the Organization for Security and Cooperation in Europe (OSCE). The set of confidence building measures established by OSCE aimed at increasing transparency, cooperation, and stability in the cyber ecosystem (Neutze and Nicholas 2015, 11).
- The Association of Southeast Asian Nation (ASEAN) Regional Forum (ARF) facilitate discussion on regional confidence building measures. The ASEAN Regional Forum established a work plan for the practical operations of cyber-confidence building measures and organized a series of workshops to bring together cybersecurity actors (Pawlak 2015, 141).
- The Tallinn Manual is an effort among NATO countries to research how international humanitarian laws apply to cyber warfare and general cyber offenses (Schmitt 2013).
- The Organization of American States (OAS) has a set of confidence building measures that aim to enhance trust within cyberspace and build a technical and legal capacity for its states. There has been consistent improvement in participating states' emergency response teams and law enforcement agencies (Pawlak 2015, 138).

2) Multi-stakeholer Approach: In the multi-stakeholder approach, corporate entities and academics establish platforms for enhancing communication and consult on best practices. As an example of private sector efforts, Microsoft proposed six cybersecurity norms, addressing the states relationship with private companies, products, and services. Microsoft also discussed the state's responsibility to assist the private sector with responding to cyberattacks and refraining from the production of cyber weapons (McKay et al 2014). The purpose of these norms is to improve defense from cyberattacks and to limit offensive operations.

Furthermore, under information-sharing confidence building measures, organizations such as the Forum of Incident Response and Security Teams (FIRST) are established as channels for the national CERT to cooperate. FIRST establishes platforms that enable international cooperation and the sharing of best practices on cyberspace security. FIRST also created an open source project aimed to provide public Application Program Interface (API) to the Computer Security Incident Response Team's database of real time updates across the globe.

C. Effect of Current CBMs

The norms and confidence building measures outlined above serve mostly as recommendations for countries to voluntarily follow. With minimal political commitment, confidence building measures do not address the technical difficulty in establishing attribution of cyberattacks, making it impossible to determine when exactly treaty agreements are breached. Furthermore, while these confidence building measures intend to prevent crisis escalation as a result of misconceptions, they do not address intentionally offensive behaviors. In spite of these limitations, confidence building measures serve as an tools for international politics by defining acceptable behavior and de-escalation mechanisms in inter-state relations (Ziolkowski 2013, 28). confidence building measures between the US and Russia and the degree to how well the two states adhere to these agreements serve as an important model for other states.

D. Unilateral Cybersecurity Policies in the US and Russia

In 2016, hacking allegations surrounding the US elections, attacks against government bodies, and theft of official government records led to tense US-Russia relations. Despite becoming a priority for both countries, there is no legal regime or governmental regulation of cyberspace (Ciglic 2016). While the US and Russia are both incentivized to reduce ICS vulnerabilities, enhance cybersecurity forces, and to avoid conflict escalation, they have

different cybersecurity policy approaches. Generally, Russia perceives cybersecurity as a part of broader information security dimension, which includes a wide range of issues such as import substitution policy, IoT, and personal data. The US approach towards cybersecurity is likely to shift and become more nationally centric and defensive. While the Obama administration had a global approach towards cybersecurity that emphasized US leadership, the Trump administration is likely to focus on defending US national interests in every domain, including cyberspace.

1) US: In 2015, the US adopted the **National Security Strategy** from a document that describes security, prosperity, American values, and international order (White House, 2015). It focuses on the US maintaining a position of leadership in international affairs; the words lead, leader, and leadership appear 94 times in the context of the country's role in the world. In the same year the Department of Defence (DoD) released **Cyber Strategy**, which listed three cyber missions: defend DoD networks and information; defend the US and its interests against cyberattacks; and provide integrated cyber capabilities to support military operations and contingency plans.

In 2016, the Obama administration released the **Cybersecurity National Action Plan (CNAP)** which includes

- the creation of a National Cybersecurity commission and IT modernization fund,
- the National Cybersecurity alliance with leading technology firms such as Microsoft, Google, Facebook along with financial services companies such as MasterCard, Visa, PayPal),
- the launch of the National Cybersecurity Awareness Campaign,
- and over $19 billion investments for cybersecurity in 2017.

Moreover, as part of CNAP, the US Cyber Command is building a Cyber Mission Force which is actively conducting cyber operations.

In 2017, the American political landscape on cybersecurity issues remains vague as President Donald J Trump transitions into power. During his campaign, President Trump announced his plans to review all US cyber defenses and vulnerabilities as well as create a Joint Task Forces throughout the US to coordinate federal, state, and local law enforcement responses to cyber threats. He has also spoken about his plans to develop American offensive cyber capabilities in order to deter attacks by both state and non-state actors. According to the the **draft Executive Order on Cybersecurity**, obtained by the Washington Post during the first month of Trumps presidency, new cybersecurity policy will include defence and enhancement of the security of the US cyberinfrastructure and capabilities along with a review of cyber vulnerabilities and cyber adversaries.

2) Russia: Russias Information Security Doctrine (Security Council of the Russian Federation 2016), published in 2016, describes the states official information and cybersecurity approach. **It defines information security as the protection of the individual, society, and state from internal and external threats in the information sphere.** The Doctrine lists a number of threats to the state, including an increased number of cyberattacks on critical information infrastructure, foreign intelligence operations against Russia, and operations that seek to destabilize the state through cybercrime. In contrast to an earlier document published in 2000, the 2016 Doctrine focuses on the military-political components that allow for Russian sovereignty in the information space. It addresses threats to the state rather than to people and businesses and includes the idea of a national Internet management system.

Another key document that will influence Russia's cybersecurity policies is the **Strategy for the Development of Information Society in Russia** for 2017-2030, published by the Russian Security Council. The document outlines the problems caused by cyberattacks on critical infrastructure and proposes solutions. The **draft law on Critical Information Infrastructure (CII)** introduces legal definition of critical infrastructure and obliges subjects of CII to share information on cyberattacks and computer incidents.

In 2016, President Putin introduced a number of specific orders which will infuence Russian cybersecurity policy. The orders imply

- use of Russian cryptographic algorithms and encryption tools by government bodies,

- prioritization of Russian software (import substitution) development of domestic technologies for the industrial Internet,
- monitoring of information threats by the Federal Security Service and other agencies,
- and regulation of personal data storage and transfer.

3) US-Russia Cybersecurity Policy Differences: Generally, Russia perceives cybersecurity as a part of broader information security dimension, which includes a wide range of issues such as import substitution policy, IoT, and personal data. The US viewed cybersecurity as a critical domain that reflects its ability to maintain international order, but that is likely to change as the Trump administration introduces new policies. The East-West Institute offers some insight into institutional differences in the way that the US and Russia handle issues related to cybersecurity. The Russian view of information security emphasizes the "holistic span of information, where cyber is one component along with others" (Godwin 2014, 11). In other words, the Kremlin considers cyber a subset of information. In the Russian language, the most similar word to the English word 'security' translates to the English word 'protection.'In contrast, the White House does not view cyber as a subset of information security, but rather its own domain, such as land, sea, or air.

Both the US and Russia agree on the emerging threats caused by cyberattacks and the need for confidence-building measures (The Russian Ministry of Foreign Affairs, 2016). In fact, in 2015 President Barack Obama and President Vladimir Putin signed an agreement on US-Russia Cooperation on Information and Communications Technology (ICT) security that was meant to facilitate a regular exchange of technical information on cybersecurity risks to critical infrastructure. However, the agreement has been largely ineffective due to geopolitical tensions caused by Russia providing asylum to Edward Snowden, the Ukrainian crisis, and western sanctions placed on Russia. In October 2016, the United States' formal accusation of Russia interfering with the election led to "normal channels of communication" being "frozen," according to Russian ambassador to the US Sergey Kislyak (Gaouette and Labott 2016). Communication withdrawal is unproductive for conflict resolution (Gulyaeva 2015) and has a potential to be catastrophic in the domain of cybersecurity, where the risk of false attribution and proportional response could compromise critical infrastructure and lead to devastating economic losses and/or civilian casualties.

The lack of uniformity in understanding cybersecurity terminology, along with the threat of communication withdrawal, begs the question: **what are the most contentious cybersecurity-related terms, or terms that prevent American and Russian officials from communicating effectively because they mean different things across cultures?** To answer this question, we turn to 'big data', or an extremely large data set that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions.

III. Identifying Discordant Cybersecurity Terms Between English and Russian Speakers

We aim to inform the problem of attribution of offensive cyber operations by adapting existing algorithms for network comparison to create a tool that identifies the most contentious cybersecurity terms between English and Russian speakers. We leverage the Global Database of Events, Language, and Tone (GDELT) to create pairs of co-occurrence networks of cybersecurity terms with related themes from content produced in the US and in Russia. The co-occurrence network pair with the most distant network similarity metric can be understood as the most contentious cybersecurity term. Upon identifying contentious cybersecurity terms, we offer a policy recommendation to tailor language to accommodate the trends found in big data.

A. Background

This section outlines the GDELT database, the concept of a co-occurence network, and network distance.

1) GDELT Database: The GDELT Database came from a desire to better understand global human society and especially the connection between communicative discourse and large-scale behavior. Its goal is to codify the entire planet

into a computable format using all available open information sources [7].

To date, the GDELT database has been used in projects that visualize the past 24 hours of global conflict and protest, perform rapid triaging and assessment of the most important influencers in an industry, topic, organization, or geographic region, and provide visibility into global trends and emerging social, political and economic risks.

This research leverages the GDELT 2.0 Global Knowledge Graph (GKG). GKG organizes a catalog of global events in over 300 categories into an annotated metadata graph over the worlds news each day. Totalling over 200 million records and growing at a rate of half a million to a million articles a day, it is perhaps the largest open data graph over global human society. It includes the GDELT Translingual feature to provide translation coverage of all monitored content in 65 languages.

2) Co-occurence Network: A co-occurence network allows us to analyze the connections among entities in a database. For a selected term, the network shows all the terms to which it is most connected to during a particular time period or in relation to a particular topic. The weight of a given edge in the network is computed as the count of that term divided by the total counts of all terms. The larger the weight, the closer the node is in the network to the selected term. Previous work involving co-occurence networks includes:

- choosing the word most typical in context using a lexical co-occurence network,
- using co-occurence network structure to synonymous gene and protein names, and
- exploring co-occurence network patterns in soil microbial communities.

3) Network Distance: **Given a set of networks of possibly different sizes, how can we efficiently provide a measure of structural similarity?** This concept of network similarity (or distance) can be approached by forming graph signatures through feature extraction. After signatures are generated, a distance function can be applied to provide meaningful network distance metrics. We rely on the work of Berlingerio et. al [9] to use an algorithm modeled after their NETSIMILE distance algorithm in order to arrive at an effective and scalable solution. NETSIMILE is attractive because it gives similarity scores that are size-invariant and because it is scalable, being linear on the number of edge for signature vector extraction.

B. Methodology

This section provides an overview of the data collection process, the formulation of co-occurrence networks, additional graph processing, and the design choices for our network distance function.

1) Data Collection Process: The GDELT 2.0 Global Knowledge Graph (GKG) is available in BigQuery, Google's enterprise data warehouse for analytics. GKG's complex multi-delimiter fields are accessible with Structured Query Language (SQL) but require advanced string manipulation features of BigQuery to generate final data results that are downloadable as a spreadsheet. BigQuery provides users with 1 free TeraByte (TB) every month to query data. Because GKG has over 2.5 TB of events generated from February 2015 to February 2016 alone, we quickly exceeded our quota during the data collection process. For this reason, only two networks are complete graphs in the data processing step of our experiment. This is acceptable because using these two networks, we can populate all other graphs to more than 50% of their full edge capacity. Because we conduct further graph processing to diversify the networks, each of our networks will consist of around the same number of significant term correlations.

We choose our terms of interest by running a query that returns the top ten themes associated with the term CYBER_ATTACK in the last year. We choose this term after contacting GDELT's creator Kalev H. Leetaru, who recommended that we use CYBER_ATTACK and provided documentation that lists all themes. GKG is still in development and there is no documentation that provides more information on a given theme (i.e. it is not completely clear what a theme titled TAX_FNCACT_WRITER means). Based on Kalev's feedback, we operate under the assumption that GDELT will provide this documentation by Spring 2017, thereby rendering the database even more useful for projects like ours.

In our query, we specify that the count variable should indicate the number of documents that mention a given theme, rather than the number

of mentions of a theme (i.e. a theme that appears 100 times in a given article is only counted as one appearance). This query provides a list of the top 10 cybersecurity-related terms across all content published in the last year, written in over 65 different languages.

2) Co-occurrence Network: Given this list of terms, we next form a query that returns a co-occurrence network. This is essential, as it is the connections among themes that are of interest, rather than the frequency of occurrence. This query compiles the list of themes from each matching article and records how many articles each pair of names appear in together. It only considers content published from November 1, 2015 to November 1, 2016 and limits the number of edges to 100 in order to maintain a computationally tractable problem. The final results of this query is a histogram of all pairs of themes and how many articles they have appeared in together. We conduct this query twice, once for content originally written in English and once for content originally written in Russian. For specifying language, we rely on the GDELT Database's Translingual feature which allows us to constrain queries to news coverage in a specific language. The result is a pair of co-occurrence networks from English content and Russian content for each cybersecurity term. See Figure 1 for visualization. Edge thickness is proportional to edge weight and node size is proportional to node degree.

After receiving the results of each query, we download the data into a CSV file. We use the python interface for Scalable Network Application Package (SNAP) to load the data into a graph data structure. SNAP is a general purpose, high performance system for analysis and manipulation of large networks.

3) Additional Graph Processing: We consider that each network has a similar structure because they are all co-occurrence networks (i.e. many edges connect a central node to every other node). In order to create distinct networks, we conduct additional graph processing. We begin by focusing on the English and Russian network pair of the term CYBER_ATTACK, the term we used for our initial query to find other cybersecurity-related terms. We make each of those networks a complete

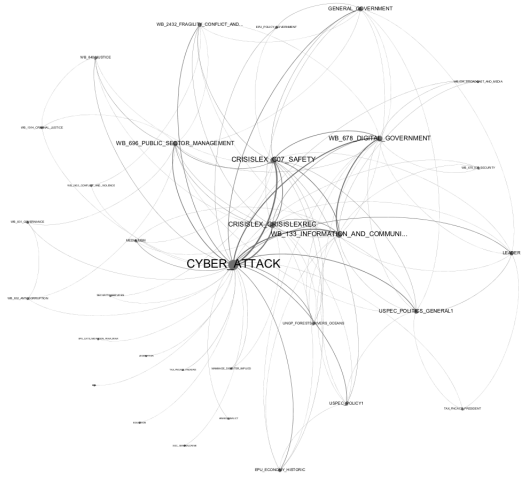


Fig. 1. Co-occurrence Network Visualization of English Term CYBER_ATTACK

graph, or a graph in which every node shares an edge with every other node. To do this, we develop a script that automates the process of iterating through each pair of nodes in the network and querying for their co-occurrence count. We then create an edge between the pair of nodes.

Next, for every other English-language co-occurrence network, we add edges between nodes if those nodes are linked by an edge in any other English network. For example, if we are looking at the English co-occurrence network of the term GENERAL_GOVERNMENT and find that nodes ARMED_CONFLICT and ELECTION exist but are not linked by an edge, we can search all of our other English networks to find the edge between these two terms because all edges between two of the same nodes are the same. If we find the edge between ARMED_CONFLICT and ELECTION in another network, we then add that edge to the GENERAL_GOVERNMENT English co-occurrence network. We do the same for the Russian networks.

This technique successfully increases the number of edges in our co-occurrence networks. For our English networks, the number of edges increases to **51%** of the edges present in the complete graph, on average. For our Russian networks, the number of edges increases to **58%** of the edges present in the complete graph, on average.

To further diversify our networks, we find the

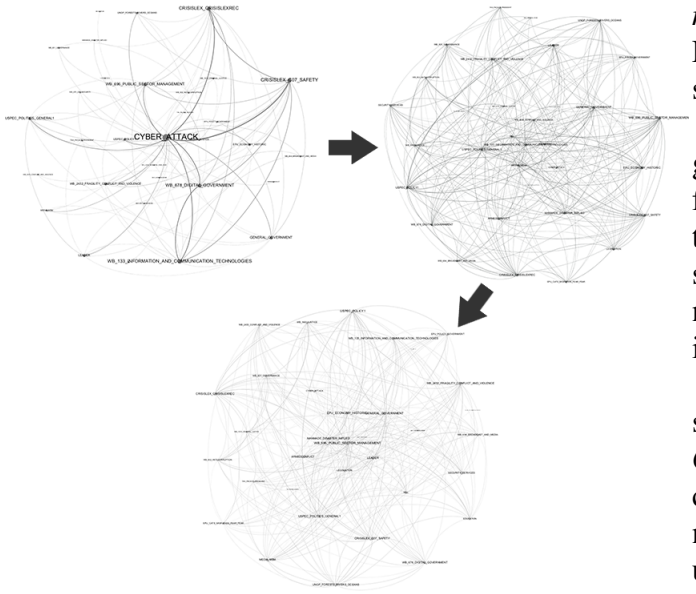


Fig. 2. Graph Processing Visualization of English Term CYBER_ATTACK

average of the edge counts for each network and use that as a threshold. For each edge in a network, if the edge count is less than our threshold, we delete it from the graph. The result is a network with edges that represent a significant correlation between nodes, or terms in our dataset. We conduct our analysis on a set of 10 structurally unique network pairs of English and Russian cybersecurity terms. See Figure 2 for a visualization of all data processing.

C. Network Distance Function

Our network distance function consists of three parts: feature extraction, feature aggregation, and network comparison.

Feature extraction. Based on the work of [9], we generate a set of structure features for each node based on its local features. Specifically, we compute the following four features: The number of neighbors, The clustering coefficient of node, The average number of a nodes two-hop away neighbors, and the average clustering coefficient.

Feature aggregation. After the feature extraction step, we have a set of feature matrices $\{F_{G_1}, F_{G_2}, ..., F_{G_k}\}$. We now generate a single "signature" vector for each graph to produce efficient and effective comparisons. We use the following five aggregators on each feature: *median, mean, standard deviation, skewness, and kurtosis.* For each node, we append these aggregators to a single list, our signature vector.

Center node features. After adding the aggregators for each feature, we compute the same features for the center node specifically and append them to the signature vector. Because our networks started off as co-occurrence networks, the center node has a significance that should be represented in our network distance function.

Network Comparison. After the aggregation step, we have a signature vector $\vec{s}_{G_j}$ for each graph $G_j \in \{G_1, G_2, ..., G_k\}$. We conduct a pairwise comparison between the English and Russian signature vectors $\vec{s}_{G_{j,ENG}}, \vec{s}_{G_{j,RUS}}$ for each graph. We use the Canberra Distance function,

$$d(\vec{s}_{G_{j,ENG}}, \vec{s}_{G_{j,RUS}}) = \sum_{i=1}^{d} \frac{|(\vec{s}_{G_{j,ENG}})_i - (\vec{s}_{G_{j,RUS}})_i|}{\vec{s}_{G_{j,ENG}})_i + (\vec{s}_{G_{j,RUS}})_i}$$

because it is discriminative, a good property for distance measure. This is because the Canberra Distance function is sensitive to small changes near zero, as it normalizes the absolute difference between graphs.

Computational Complexity. As Belangario et al. (Berlangario et al. 2012) demonstrates, the runtime complexity of this algorithm is

$$O(\sum_{j=1}^{k}(fn_j + fn_j log(n_j)))$$

where k is the number of graphs, n_j the number of nodes, and f the number of structural features extracted. In real world networks,

$$f << n_j << m_j$$
$$n_j log(n_j) \approx m_j$$

where m_j is the number of edges. In other words, the algorithm is linear on the number of edges when used on our GDELT database.

D. Results

This section provides the empirical results and evaluations of our experiments.

Cybersecurity Terms. We began by finding the top ten cybersecurity terms across all languages. They are enumerated in Table 1 along with their counts.

Algorithm 1 Generate Signature Vectors

procedure
$\{F_{G_1}, F_{G_2}, ..., F_{G_k}\}$ = *GetFeatures*
for all $x \in \{F_{G_1}, F_{G_2}, ..., F_{G_k}\}$
    centerNode = getCenterNode(x)
    $s_{G_x} = []$
    for all *feat*$\in F_{G_x}$
        $s_{G_x} \cup$ {*median(feat), mean(feat), stdev(feat), skewness(feat), kurtosis(feat)*}
    end for
    $s_{G_x} \cup$ *getNeighborNo(centerNode)*
    $s_{G_x} \cup$ *getNodeCC(centerNode)*
    $s_{G_x} \cup$ *getAvgTwoHop(centerNode)*
    $s_{G_x} \cup$ *getAvgCC(centerNode)*
end for
return $\{s_{G_1}, s_{G_2}, ..., s_{G_k}\}$

TABLE I
CYBERSECURITY TERMS

TERM	COUNT
CYBER_ATTACK	1888964
CRISISLEX_C07_SAFETY	1251265
CRISISLEX_CRISISLEXREC	1070679
WB_133_INFORMATION_AND_COM.	1022871
WB_678_DIGITAL_GOV	991150
WB_696_PUBLIC_SECTOR_MAN	776948
USPEC_POLITICS_GENERAL1	733939
GENERAL_GOVERNMENT	689712
WB_2432_FRAGILITY_CONFLICT	687723
USPEC_POLICY1	684035

Terms that include information and communication, digital government, public sector management, general politics, general government, conflict and violence, and policy are expected, as cybersecurity policy has become a significant area of concern for both governments in the past year.

Network Distance. The network distance results of English-Russian word pairs associated with the cybersecurity-related terms can be seen in Figure 3. The median distance score is 5 while the mean distance score is 4.64. Interestingly, rather than being a discordant word, CYBER_ATTACK is well below our median, with a score of 0.54. The term CRISISLEX_C07_SAFETY has a negative distance metric. We conclude that these two terms must have especially similar network structures between the English and Russian translations of

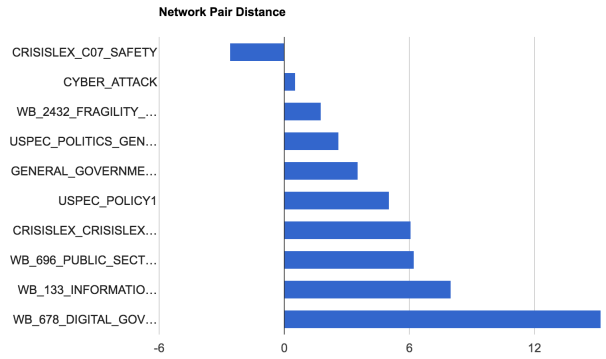


Fig. 3. Network distances of English-Russian Cybersecurity term Network Pairs

the term. In other words, these terms are contextualized similarly in English and Russian content. Terms that indicate policy, public sector management, information, and communication technologies are above average in discordance. This suggests a fundamental difference in how English and Russian speakers understand the role of government in regulating information and communication technologies.

Modularity. US and Russian policymakers will benefit from further analysis on what exactly contributes to the different viewpoints suggested above. Analyzing community structure within our networks has the potential to inform policymakers on the themes that drive discordance. We explore this by running a **community detection algorithm** (Vincent et al. 2008) on our most discordant term, WB_678_DIGITAL_GOVERNMENT. For our parameters, we choose to consider edge weight and set the resolution to 0.5, where 1 represents an algorithm that prefers large communities and 0 represents an algorithm that prefers small communities. For our English network, the result is **4 communities**. For our Russian network, the result is **6 communities**. A conclusion we draw from our modularity findings is that more themes go into the concept of digital government in Russian content than in English content.

Network Distance Features. Finally, we explore our network distance features and hypothesize which features drive network distance. For each of our signature vectors, we compute the Canberra Distance function of each of our feature values and analyze which features contribute the most

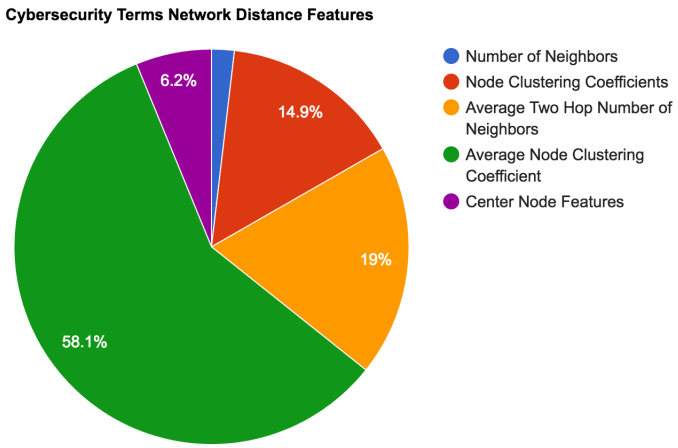


Fig. 4.

to the total distance. Figure 4 demonstrates that the feature Average Node Clustering Coefficient contributed the most while the feature Number of Neighbors contributed the least. It is expected that center node features do not contribute much to the distance score because additional graph processing reduced the degree of the center node.

The significance of the feature Average Node Clustering Coefficient may stem from its interpretation as a global clustering coefficient. A graph is considered small-world if its average clustering coefficient is significantly higher than that of a random graph. We hypothesize that a high average clustering coefficient and a high average two hop number of neighbors metric indicate a network is a small-world network, meaning that all nodes have a low degree of separation. This suggests that between English and Russian translations of a term, one network is highly connected while the other is not. This may be because a word is either highly contextualized in content, with many links to other words, or not contextualized very much. In other words, either a word means something very specific to an English or Russian speaker, or it is not well-defined at all.

IV. Conclusion

Our policy analysis demonstrates a clear danger in false attribution because of the challenges inherent to cybersecurity policy and the weak communication channels that exist between the US and Russia. Without a willingness to communicate, existing American and Russia organizational structures meant to avoid conflict escalation are rendered ineffective. As both states unilaterally implement nationally-oriented laws, there exists the heightened threat of proportional response and conflict escalation.

With no clear international legal code or lead cybersecurity organization in place, a policy recommendation that centers around strengthening American and Russian communication strategies is needed now more than ever.

Our recommendation to policymakers is, when dealing with discordant terms, to attempt to understand the exact definition of the term from both perspectives at the start of negotiations in order to avoid confusion or conflict escalation later. Our empirical analysis findings include the following:

- Terms that indicate policy, public sector management, information, and communication technologies are highly discordant. This affirms a difference outlined in Section II: there exists a fundamental difference in how English and Russian speakers understand the role of government in regulating information and communication technologies. While cybersecurity falls under the umbrella of information security in Russia, in the US it is its own military domain and perhaps less likely to touch the lives of civilians than in Russia.
- Russian networks demonstrated a high degree of modularity, or community structure. We conclude that more themes go into the concept of digital government in a Russian content than in English content. Practically, this may mean that to a Russia speaker, an understanding of cybersecurity is more nuanced that to an English speaker.
- Due to the significance of the Average Node Clustering Coefficient as a driver of network distance, we conclude that for both English and Russian speakers, cybersecurity terms are either highly contextualized or not well-defined. This suggests specific definitions for each term in both cultures, which could promote conflict escalation if these definitions are not understood by both sides at the start of negotiations

While these suggestions are useful as a foundation for communication that informs each side

without bias, further work can be done on contextualizing cybersecurity terms between English and Russian speakers. Future work in this area involves expanding our signature vector to include more features and expanding our exploration of community structure.

Ultimately, the future of global cybersecurity depends on policies made now. The US and Russia serve as an important model for other countries. It is critical that the foundation of negotiations be a strong communication channel and an understanding of different cybersecurity contexutalizations across cultures, so that both governments can operate with a high likelihood of cooperation.

V. ACKNOWLEDGEMENTS

The authors would like to thank the Stanford US Russia Program for providing a platform for international collaboration and Dr. Tatiana Tropina for her guidance in the development of this paper.

REFERENCES

[1] Rainie, Lee, Janna Anderson, and Jennifer Connolly. "Cyber Attacks Likely to Increase." Pew Research Center: Internet, Science & Tech. October 29, 2014. Accessed February 22, 2017. http://www.pewinternet.org/2014/10/29/cyber-attacks-likely-to-increase/.

[2] Owens, William A., Kenneth W. Dam, and Herbert S. Lin, eds. Technology, policy, law, and ethics regarding US acquisition and use of cyberattack capabilities. National Academies Press, 2009.

[3] Andreeva, Oxana, and Sergey Gordeychik."Industrial Control Systems And Their Online Availability." (2016) https://kasperskycontenthub.com/securelist/files/2016/07/KL_REPORT_ICS_Availability_Statistics.pdf

[4] Stouffer, Keith, Victoria Pillitteri, Suzanne Lightman, Marshall Abrams, and Adam Hahn. "Guide to Industrial Control Systems (ICS) Security." 2015. doi:10.6028/nist.sp.800-82r2

[5] McMillen, Dave Attacks Targeting Industrial Control Systems (ICS) Up 110 Percent. IBM. 2016. https://securityintelligence.com/attacks-targeting-industrial-control-systems-ics-up-110-percent/

[6] Jasper, Scott. Deterring Malicious Behavior in Cyberspace. Strategic Studies Quarterly. 2015. Air University Maxwell Air Force Base.

[7] Levy, Jack S. "The Offensive/Defensive Balance of Military Technology: A Theoretical and Historical Analysis." International Studies Quarterly 28, no. 2. 1984: 219-38. doi:10.2307/2600696.

[8] Cirenza, Patrick. "An Evaluation of the Analogy Between Nuclear and Cyber Deterrence." (2015). Accessed February 21, 2017 https://cisac.fsi.stanford.edu/sites/default/files/cirenza_final_thesis_2015.pdf

[9] Neutze, Jan, and J. Paul Nicholas. "Cyber Insecurity: Competition, Conflict, and Innovation Demand Effective Cyber Security Norms." Georgetown Journal of International Affairs, 2013, 3-15. http://www.jstor.org/stable/43134318.

[10] Pawlak, Patrick. "Confidence-Building Measures in Cyberspace: Current Debates and Trends." In International Cyber Norms: Legal, Policy & Industry Perspectives edited by Anna-Maria Osula and Henry Rigas, 129-153. Tallinn: NATO CCD COE Publication,

[11] Schmitt, Michael N. Tallinn Manual on the International Law Applicable to Cyber Warfare. New York: Cambridge University Press, 2013

[12] Ziolkowski, Katharina. "Confidence Building Measures for CyberspaceLegal Implications." Tallinn: NATO CCD COE Publication (2013).

[13] Berlingerio, Michele, Danai Koutra, Tina Eliassi-Rad, and Christos Faloutsos. "NetSimile: a scalable approach to size-independent network similarity." arXiv preprint arXiv:1209.2684 (2012). Browse on GitHub CrisisLexLexicon-v1.0.zip (2.9 KB)

[14] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, Fast unfolding of communities in large networks, in Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P1000

[15] United Nations, General Assembly, Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, A/70/174 (22 July 2015), available from undocs.org/A/70/174.

[16] Ciglic Kaja, seminar, December 9, 2016.

[17] Security Council of the Russian Federation. Information Security Doctrine of the Russian Federation. Moscow: 2016.

[18] Security Council of the Russian Federation. Draft of Strategy of Information Society Development of the Russian Federation. Moscow: 2016.

[19] State Duma of the Russian Federation. Draft law on Critical Information Infrastructure (CII). Moscow: 2016.

[20] The President of the Russian Federation. Orders on ensuring the development and implementation of a set of measures necessary for the transition of authorities on the use of Russian cryptographic algorithms and encryption tools. Moscow: kremlin.ru, 2016. http://kremlin.ru/acts/assignments/orders/52536

[21] The President of the Russian Federation. Orders following the first Russian Forum Internet economy. Moscow: kremlin.ru, 2016. http://www.kremlin.ru/acts/assignments/orders/51235

[22] "Obama's Last National Security Strategy." Foreign Affairs. March 02, 2015. Accessed February 23, 2017. https://www.foreignaffairs.com/articles/united-states/2015-03-02/obamas-last-national-security-strategy.

[23] US President. Draft Executive Order. Strengthening US Cyber Security and Capabilities. https://assets.documentcloud.org/documents/3424611/Read-the-Trump-administration-s-draft-of-the.pdf

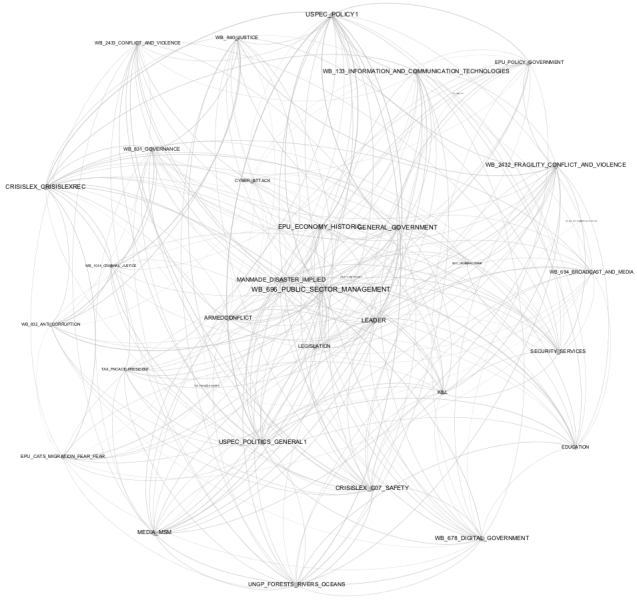
[24] Godwin III, J. B., Kulpin, A., Rauscher, K. F., and Yaschenko, V., editors (2014). Critical Terminology Foundations 2. East-West Institute and the Information Security Institute at the Moscow State University. https://www.files.ethz.ch/isn/178418/terminology2.pdf

APPENDIX I
CODE

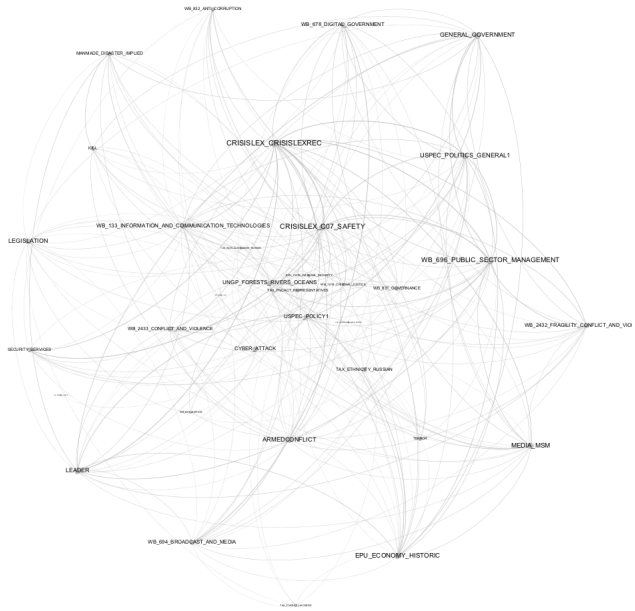
Implementation can be found at `https://github.com/ashemag/CybersecurityContextUSRussia`

APPENDIX II

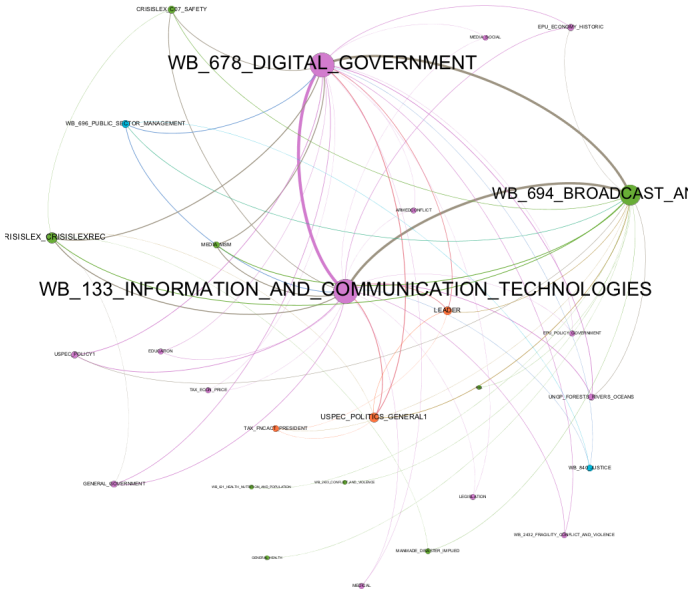
APPENDIX III
FINAL NETWORK FOR ENGLISH TERM CYBER_ATTACK



APPENDIX IV
FINAL NETWORK FOR RUSSIAN TERM CYBER_ATTACK



APPENDIX V
COMMUNITY DETECTION NETWORK FOR ENGLISH TERM WB_678_DIGITAL_GOVERNMENT



APPENDIX VI
COMMUNITY DETECTION NETWORK FOR RUSSIAN TERM WB_678_DIGITAL_GOVERNMENT