Junk News on Military Affairs and National Security: Social Media Disinformation Campaigns Against US Military Personnel and Veterans

COMPROP ONLINE SUPPLEMENT TO DATA MEMO 2017.9 / 09 OCTOBER 2017

John D. Gallacher
Oxford University
john.gallacher@cybersecurity.ox.ac.uk
@john_gallacher1

Vlad Barash
Graphika
vlad.barash@graphika.com

Philip N. Howard
Oxford University
philip.howard@oii.ox.ac.uk
@pnhoward

John Kelly
Graphika
john.kelly@graphika.com

Appendix 1. The Audience for Veterans Operations and Related Content on Twitter

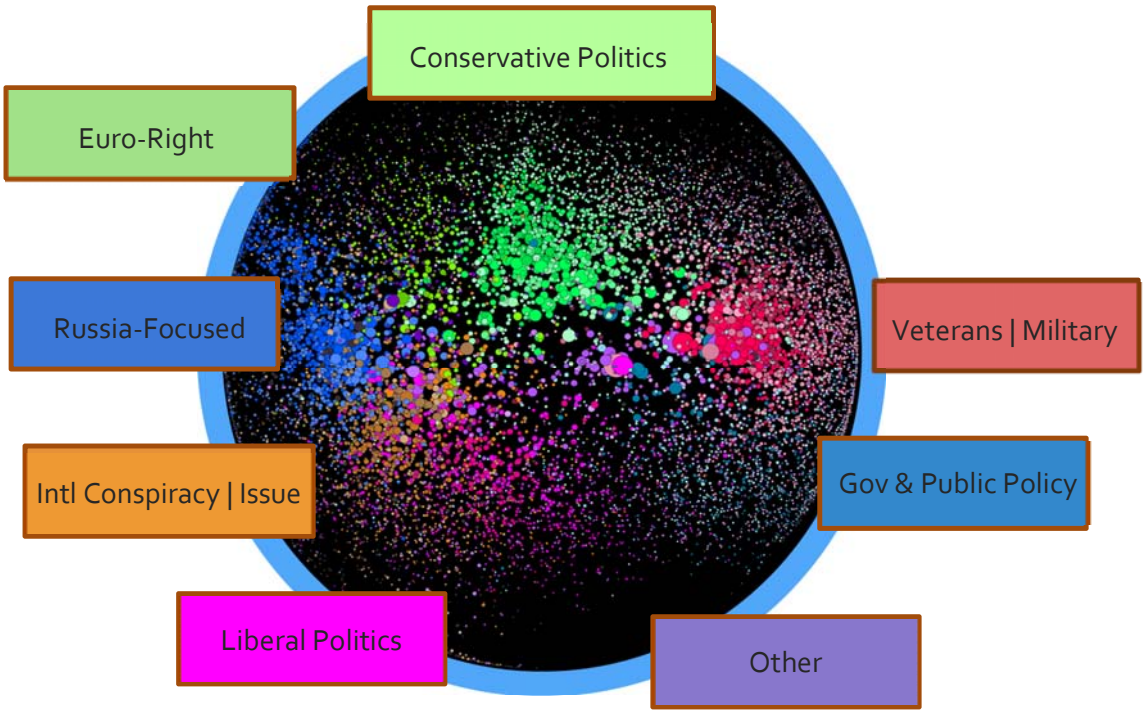

Figure 1. Full visualisation of the audience for Veterans Operations and Related Content on Twitter Authors' calculations from data sampled 02/4/-02/5/ 2017

Each node in this network represents an account on Twitter. Each node belongs to both a broad group and a smaller segment within that group. The size of the node is proportional to the number of other map nodes that follow it on Twitter. The colour of the node is based on its parent segment.

A segment is a collection of nodes with a shared pattern of interest while a group is a collection of segments that are geographically, culturally, or socially similar.

The nodes are placed within the map using a Fruchterman-Reingold visualization algorithm. This works to place nodes into the map according to two principles: first, a "centrifugal force" acts upon each node to push it to the edge of the canvas; second, a "cohesive force" acts upon every connected pair of nodes to push them closer together.

Full list of Groups and Segments for the Twitter Map

Group	Segment	Group	Segment	Group	Segment
Russia-Focus	Pro Putin Trolls / Pols	Conservative Politics	Conservative Pundits / Celebs	Gov & Public Policy	Public Health
	Foreign Policy Journos / MENA		Trumpista		Beltway Polit /Congress
	Pro Putin Russian Trolls Abroad		True American Patriotism		Tech and Finance News
	Pro Assad / Russia / Trump		Real Donald Trump		Nonprofit / Eco / Education
	Pro Putin Russians / Ukraine		Conservative Pundit / Fox		US Gov / Emergency Response
Intl Conspiracy / Issue	Anti-NWO		Tea Party / Guns	Other	Pop Culture
	Pro-Palestine		Pro-Trump Core		Pop Culture
	US Libertarian		Constitutional Conservatives		Celeb / Wrestling Focus
	Intl RT and Wikileaks	Liberal Politics	Pro-Bernie / Resist		SMM Inspiration
Veterans / Military	US Military 2		UK Left		Central / Eastern Europe Politics
	US Military / Navy / Marines		Progressives		Foreign Policy Intl / US
	Defence Industry		US Liberals		SMM Motivation
	Conservative / Veteran 1		Prog Journo / Activism		
	Army / National Guard	Euro-Right	UKIP		
	Veterans		White Identity		
	Veterans				
	Veteran Support				
	Military Families				

Appendix 2. Audience for Veterans Operations and Related Content on Facebook

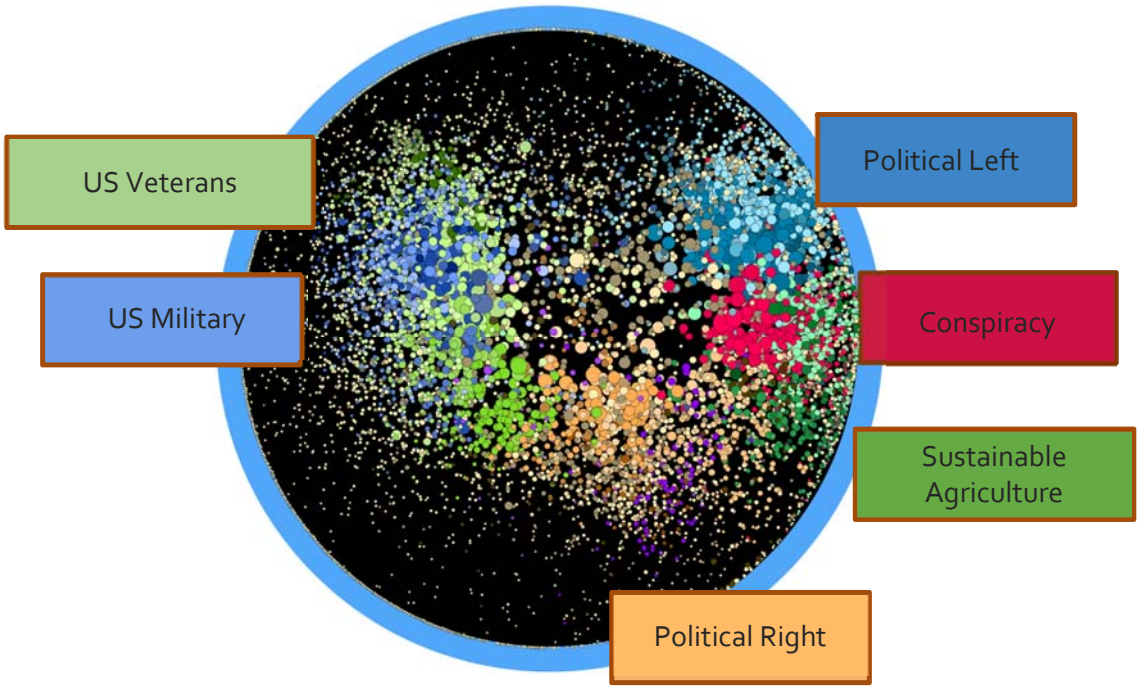

Figure 2. Full visualisation of the audience for Veterans Operations and Related Content on Facebook
Authors' calculations from data sampled 26/5/-25/6/ 2017

Each node in this network represents a public page on Facebook. The size of the node corresponds to the number of other nodes in the map that like the page on Facebook. Each node belongs to both a broad group and a smaller segment within that group. A segment is a collection of nodes with a shared pattern of interest while a group is a collection of segments that are geographically, culturally, or socially similar

Again, a Fruchterman-Reingold visualization algorithm is used to place nodes within the map.

Full list of Groups and Segments for the Facebook Map

Group	Segments	Group	Segments	Group	Segments
US Veterans	Veterans Networks/ Disability	Political Right	Libertarian / Youth	Conspiracy	Conspiracy / RT
	Veteran Support|/ Families		Libertarian Institutions		Truth / Truthers
	Veterans Networks		House Republicans		Far Right / Conspiracy
	US Military / Veteran Support		Conservative Media	Sustainable Agriculture	Organic / Sustainable Ag
	Military Gear / Weapons		Conservative Pundits		Health / Nutrition
	US VA		Hard Conservative		Anti-GMO
	VA Hospitals		Prepper / Survivalist		Small Farms / Canning
	Veterans Support		Conservative and Pro-Israel		Natural Living / Organic
	US Mil Community		Guns	Mental Health	Mental Health
	American Legion		Libertarian and End Fed Reserve		Life Coach and Meditation
US Military	US Army / Armed Forces		US Far Right and Anti-Immigrant		Sobriety and Addiction Recovery
	US Military		Conservative/ Townhall	Other	News and US Conservative
	US Navy		Hard Right / Pro-Military		Animal Lovers and Rescue
	US Military Europe / Africa	Political Left	Conservation		Anarchist
	US Army / National Guard		Womens Issues|Intl		Syria and Assad
	US Coast Guard		Western Liberal Media		
	Navy Seals / Special Ops		Labor Rights|Unions		
	US Air Force		Intl Occupy		
	US Marines		Progressive Dems		
	National Guard		US Occupy		
	US Forces / Korea		Intl Direct Democracy|Anon		
			Occupy		
			Occupy|Economic Inequality		

Additional Methodological Descriptions

Appendix 3. Heterophily Index

For every pairing of groups within a network map, a value of heterophily can be calculated. This is a measure of the level of connection between the groups. In order to determine this a ratio is calculated of the actual ties between two groups compared to the expected ties between the groups if all the accounts in the map were evenly distributed.

The natural log of these ratios is then taken, along with a zero correction to create a balanced index and ensure that all values are displayed in a positive form.

$$Ratio\ of\ \ Ratios_T = \frac{\frac{Connections_{\text{pairing}}}{\sum_{\text{all pairings}} Connections}}{\frac{Connections_{\text{pairing}}}{\sum_{\text{all pairings}} Connections}}$$

Expression A: Ratio of Two Ratios

This heterophily index is therefore created through a ratio of two ratios. The ratio of these two ratios reveals whether two nodes have about the proportion of links they should have given its size. This is displayed in Expression A, where a pairing of groups is calculated as having a measure of connections in balance with its share of all the connections.

Half the distribution of possible values from this ratio of ratios ranges from 0 to 1 (a disproportionately small share of connections in a group given its size) and the other half ranges from 1 to +infinity (a disproportionately large share of connections in a group given its size). However, by taking the natural log of the ratio of ratios the index will become more balanced: from -infinity to 0 becomes less than proportionate share, and from 0 to +infinity becomes more than proportionate share.

For example, take a three-group network (A, B and C). If nodes in group A have a total of ten connections, and there are ten nodes in each group, then the expected connections between A and B will be 3.33. If, in reality, the nodes in group A actually have all ten connections to nodes in group B then this connection is stronger than expected. The heterophily score for groups A and B = 10/3.33 = 3.0. The natural log of this is then taken along with a zero correction across the range of heterophily values.

A greater heterophily index indicates a denser pattern of connections between the two groups. It is important to note however that these scores indicate only first order connections, not second or third order connections.

Appendix 4. Clustering for groups and Segments

In order to generate segments and groups for each map it is necessary to employ a clustering algorithm.

This involves first building a bipartite graph between nodes in the map and the rest of the social medium in question. This bipartite graph provides a structural similarity metric between nodes in the map.

This was then used in combination with a hierarchical agglomerative clustering algorithm in order to segment a map into distinct communities. This is a 'bottom up' approach whereby each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.

Twitter maps are clustered based on follower relationships, since mentions relationships have been shown to overemphasize the news cycle and salient external events. Facebook networks are clustered based on page likes.

Appendix 5. K-core reduction

To identify and map the 'discussion core' of the most active, connected, and influential users, we performed a k-core reduction to reduce the total collected set of Twitter users from the initial data collection into a set of well-connected accounts. This produces a maximally connected subgraph of active nodes with degree of connection at least *'k'*.

This degree of connection, k, can be thought of as the number of links between each node in the graph. For example, selecting a k value of 0 for the reduction not remove any nodes from the graph, since each node must have 0 connections or greater. Selecting a k value of 1 would remove all of the nodes that have no connections to other nodes in the graph. Selecting a k value of 2 would remove all nodes with fewer than 2 connections, and so on.

A value of k was selected such that the k-core consisted of 12,413 users. This value was found to be a sufficiently large group to represent the major sets of highly active users, but not so large as to make clustering and visualization impractical.