

**Applied Physics Laboratory
University of Washington**

FINAL TECHNICAL REPORT

**ONR Grant N00014-16-1-2919
Factional Networks in Fragmented Conflicts**

Principal Investigator

Dr. Michael Gabbay
Applied Physics Laboratory
University of Washington
1013 NE 40th Street
Seattle, WA 98105

Period of Performance: September 1, 2016 – November 30, 2021

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY)		2. REPORT TYPE Final Technical		3. DATES COVERED (From - To) 9/01/2016 - 11/30/2021	
4. TITLE AND SUBTITLE Factional Networks in Fragmented Conflicts				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER N00014-16-1-2919	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Michael Gabbay				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Washington - Applied Physics Laboratory 4333 Brooklyn Avenue NE Seattle, WA 98105-6613				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of Naval Research/Code 0341 875 North Randolph Street Arlington, VA 22203-1995				10. SPONSOR/MONITOR'S ACRONYM(S) ONR	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT: Distribution Statement A: Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT This report describes research exploring the use of network methods to analyze the interactions of militant groups in insurgencies and civil wars. Network and attribute data for militants were collected using manual and automated methods; theories of militant factional behaviors were developed and tested empirically; novel methods of analyzing militant network structure were developed; and new results on the fundamental structure and dynamics of signed networks were obtained.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Michael Gabbay
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (include area code) (206) 543-1300

Table of Contents

1	Scope	3
2	Accomplishments	3
3	Summary of Research.....	4
3.1	Syrian Militant Network Data	4
3.2	Ukrainian Militant Network Data	5
3.3	Structural Analysis Methods	7
3.4	Drivers of Militant Cooperation and Conflict	9
3.5	Modeling Signed Network Structure and Dynamics.....	11
3.6	Tracking and Anticipating Militant Conflict.....	13
4	Transition Activity	16
5	Training	17
6	Publications and Presentations	17
6.1	Publications	17
6.2	Presentations.....	18

Appendices

Appendix 1: Fratricide in Rebel Movements: A Network Analysis of Syrian Militant Infighting

Appendix 2: Networks of Cooperation: Rebel Alliances in Fragmented Civil Wars

Appendix 3: Integrating Computational Modeling and Experiments: Toward a More Unified Theory of Social Influence

Appendix 4: Community Detectability and Structural Balance Dynamics in Signed Networks

Appendix 5: Militants and Mixed Messages: The Effect of Sponsor Ambiguity on Insurgent Ideology and Networks

1 Scope

This report describes research conducted by the Applied Physics Laboratory of the University of Washington (APL-UW) for the Office of Naval Research (ONR) under grant N00014-16-1-2919. The research performed under this grant explored the use of network methods to analyze the interactions of militant groups in insurgencies and civil wars. Network and attribute data for militants were collected using manual and automated methods; theories of militant factional behaviors were developed and tested empirically; novel methods of analyzing militant network structure were developed; and new results on the fundamental structure and dynamics of signed networks were obtained. The principal investigator (PI) of this effort research was Dr. Michael Gabbay.

2 Accomplishments

The following are significant accomplishments of this effort:

1. Collected high resolution data of militant group cooperation and conflict in Syrian Civil War and the separatist insurgency in Eastern Ukraine (based on US government translated documents).
2. Developed a pipeline for collecting Ukrainian militant network and ideology data directly from social media (VKontakte) integrating machine learning and manual content analysis.
3. Developed and refined multiple network analysis methods, particularly for signed networks.
4. Published the first statistical network analyses of how ideology, power, and state sponsorship affect militant cooperation and conflict using the Syrian data. Received the Journal of Peace Research 2019 best paper award.
5. Developed a theory of how ideology can shape militant networks even for the hard case of a single state sponsor and empirically tested the theory for the case of Ukrainian separatists and Russia.
6. Conducted the first signed network analysis of militants in a civil war using the Ukraine data. Validated that ideology can shape militant networks for the single state sponsor case.
7. Demonstrated the existence of phase transitions in signed networks related to the detectability of community structure and prosocial behavior.
8. Investigated how signed networks evolve under structural balance dynamics. Showed that the above phase transitions induce sharp boundaries between final state regimes of identity-based factional polarization, identity-mixed factional polarization, and universal harmony.
9. Developed a novel measure of polarization in signed networks that integrates community structure and structural balance.
10. Using the Syrian data, empirically showed how this “eigenvector polarization” metric tracks the level of conflict and detects the onset of systemic war within a network.
11. Published a framework for integrating computational and experimental studies of social influence.

3 Summary of Research

In this section, we summarize key elements of this research effort. Reference numbers refer to documents listed in Section 6.1.

3.1 Syrian Militant Network Data

We collected an extensive dataset on infighting between Syrian militant groups using manual coding of US Government documents of translated militant statements; the data includes 508 episodes of infighting yielding 631 conflict dyads between 32 groups. Another dataset was developed on 14 major alliances (coalitions, mergers, and operation rooms) between September 2012 and January 2017. Ideology, group size, state sponsorship, and geographic zones of operation were also coded for 43 Syrian groups. Our previously collected data on joint operations of Syrian militant groups was updated to run through June 2015 and consists of 696 joint operations and more than 900 ties between 220 groups. When first constructed, this data represented the most extensive dataset of conflictual and cooperative ties between militant groups in a single movement. A description of the procedures used to construct the data are described in Refs. [1] and [2]. Figure 1 shows the combined infighting and joint operations networks.

The Syria militant data initially collected under this effort was extended to include more groups, extend the time period through 2018, and include data on attacks by the Syrian regime. This work was done in collaboration with a grant funded by the Air Force (PI: Dr. Emily Gade). A paper describing the dataset is in preparation.

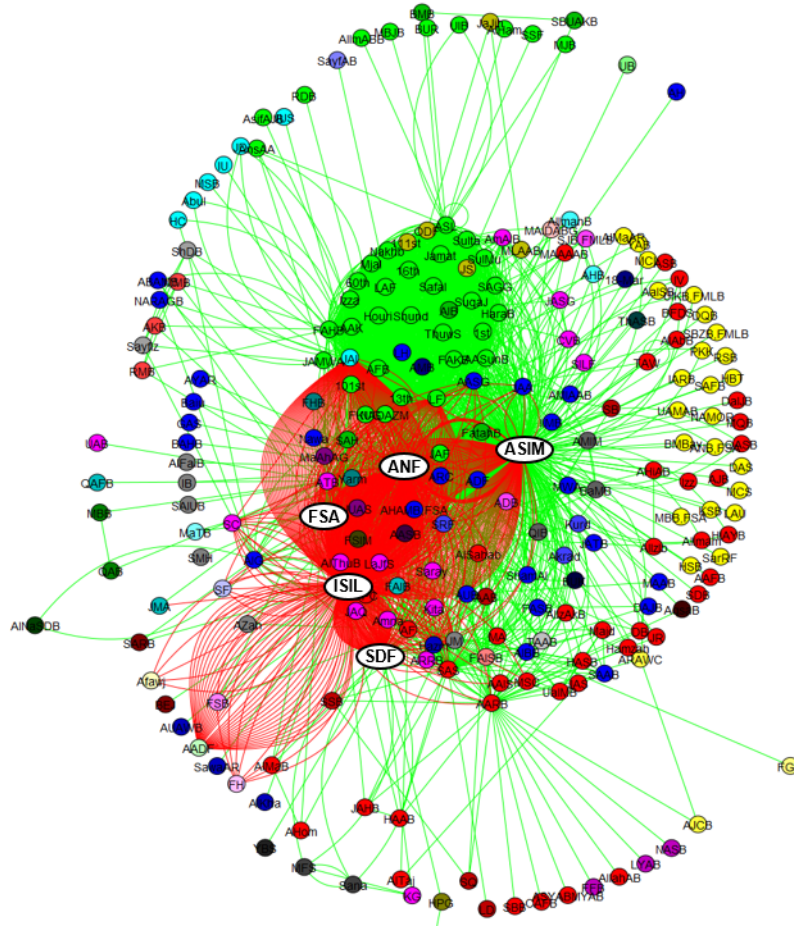


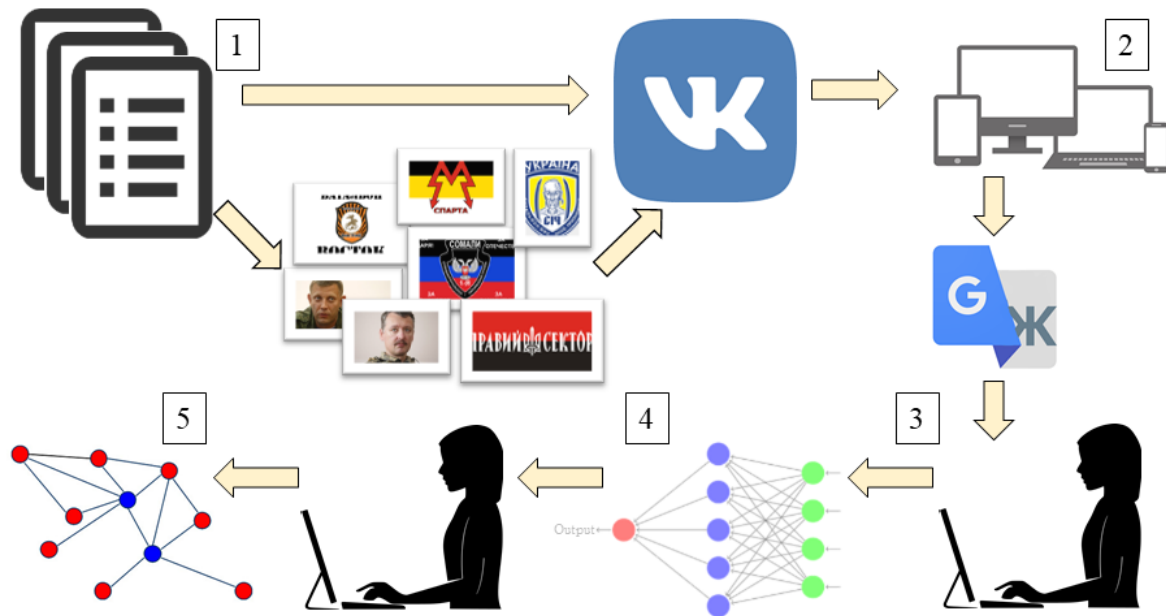
Figure 1. Network of Syrian militant groups showing cooperation and conflict. Joint operations between groups are shown in green and infighting episodes are shown as red ties. Nodes are colored by communities detected by the Louvain algorithm for community detection in signed networks. ISIL: Islamic State, ANF: Al Nusra Front, ASIM: Ahrar al-Sham, FSA: Free Syrian Army, SDF: Syrian Democratic Forces (Kurds)

3.2 Ukrainian Militant Network Data

We collected data on cooperation and conflict between militants in Eastern Ukraine using manual coding of US Government translations of militant rhetoric. We collected data on both militant groups and leaders. The complete data includes 467 events (216 cooperative, 251 conflictual) between 87 actors (out of 258 being tracked). Actor ideology along the three dimensions of militant ideology originally developed for Syria (conflict frame, territorial aspiration, ideal polity) was also coded. Our subsequent analysis of the data for separatist militants focused on their leaders as they were more likely to make statements with ideological relevance. This data consisted of 28 leaders and 171 interaction events (86 cooperative ties and 85 conflictual). Reference [5] describes the data in more detail.

Due to the shutdown of the service providing access to the government translations, we shifted to collecting Ukrainian militant data directly from social media. The method involves downloading

and pre-processing posts from conflict-related accounts on VKontakte. The posts are automatically translated and evaluated for relevance. The relevance evaluation is done using machine learning employing natural language processing (BERT) and a classification algorithm, trained on manually-labeled posts. Coding rules were developed and undergraduate RAs conducted manual coding of militant interaction events and ideology in the social media data. The processing pipeline is shown in Figure 2. A signed network generated from the data which contains cooperative and conflictual events between actors on both the separatist and Ukrainian government sides is shown in Figure 3.



1. Search of OSE documents, journal articles, monographs, and Wikipedia pages identifies list of key militant actors, as well as VK pages relevant to conflict. Additional searching done to match actors to VK accounts.
2. Data accessed programmatically via VK API, then translated using Google Translate API.
3. Human coders label posts from militant accounts as relevant or not relevant.
4. Natural language processing (BERT + classification model) used to predict whether each post in corpus is relevant.
5. Human coders read relevant posts to generate networks of insurgent cooperation & conflict, along with insurgent ideology.

Figure 2. Social media processing pipeline for Ukrainian militant events and ideology.

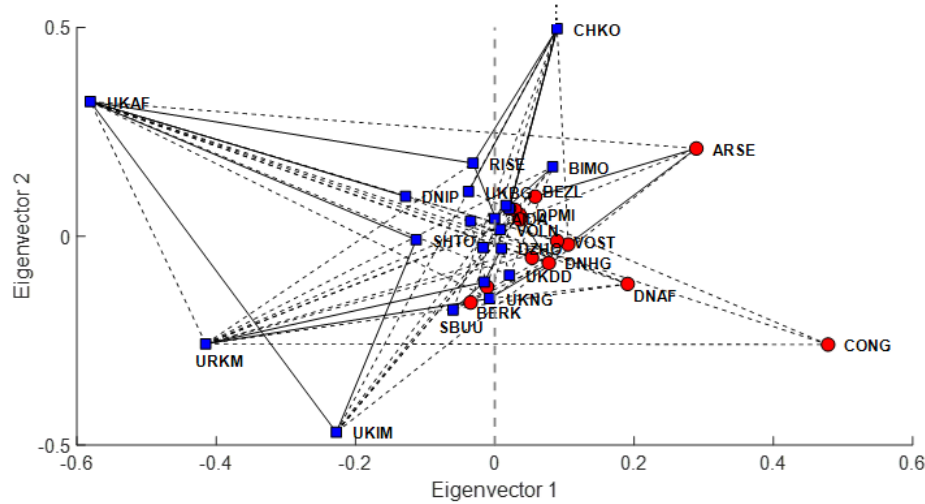


Figure 3. Signed network of militant and other armed actors in Eastern Ukraine from social media. Red circles are separatist actors; blue squares are actors affiliated with the pro-Ukrainian side. Solid and dashed lines indicate cooperative and conflictual events respectively.

3.3 Structural Analysis Methods

We employed a number of network analytic tools in our analysis of militant factional behavior in order to assess the robustness of theorized relationships between network structure and node variables such as ideology and power. In particular, whether tie formation is driven by attribute similarity (homophily) or dissimilarity (heterophily). Our work on the Syrian insurgency focused on treating the cooperation and conflict networks separately. We employed the following methods that we developed in Matlab for analyzing unsigned networks:

- (1) Modularity spectrum representation of community structure. This method tests for significant correlation of the variable of interest with network structure as represented by modularity matrix eigenvectors with large magnitude eigenvalues.
- (2) Null simulation of the assortativity metric of homophily/heterophily. This method compares the assortativity of the observed network with the mean value obtained from a Monte Carlo null simulation in which the tie formation process is driven by node degrees only (and so does not involve the variable of interest). The distribution obtained from the simulation enables the statistical significance of the difference between the observed and simulated means to be estimated. If the observed metric is significantly higher (lower) than the null mean, homophily (heterophily) is observed.
- (3) Direct simulation of tie formation including the variable of interest. This method seeks to find the interaction length which minimizes the squared error between the observed and simulated networks. If a well-defined minimum can be found, then a relation between network structure and the variable of interest is indicated.

These methods are described in Refs. [1] and [2]. In addition, we also employed network regression methods based on latent space estimation (the R-based AME package) and exponential random

graph modeling (ERGM), which enable the simultaneous testing of the relative effects of multiple variables.

Our collection of both cooperative and conflictual interactions enabled us to construct signed networks in which positive ties represent cooperation and negative ties represent conflict. To our knowledge, our work represents the first signed network analysis of militants in a civil war. We extended to the signed network case the above modularity spectrum method for determining if community structure is related to a node variable (Figure 4). This method can be used directly on the signed network adjacency matrix in addition to the modularity matrix because the first eigenvector of signed networks can represent community structure, in contrast to unsigned networks.

We employed a different assortativity null simulation than the one for unsigned networks. The method involves randomly permuting the signs of the ties present in the observed network in order to generate a null distribution of the (signed network) assortativity, which can then be compared with the observed assortativity value (see Ref. [5]). This method tests for homophily conditional upon interaction in that it only involve dyads that are connected by a tie, either positive or negative. Accordingly, this test illustrates an advantage of the signed network approach. In unsigned networks, a finding of homophily could be possibly due to a biased selection pool of like partners rather than a true preference for linking with similar others (unless non-network data is available to show that the pool is not biased). The fact that the sign swapping test is conditional upon interaction evades this criticism.

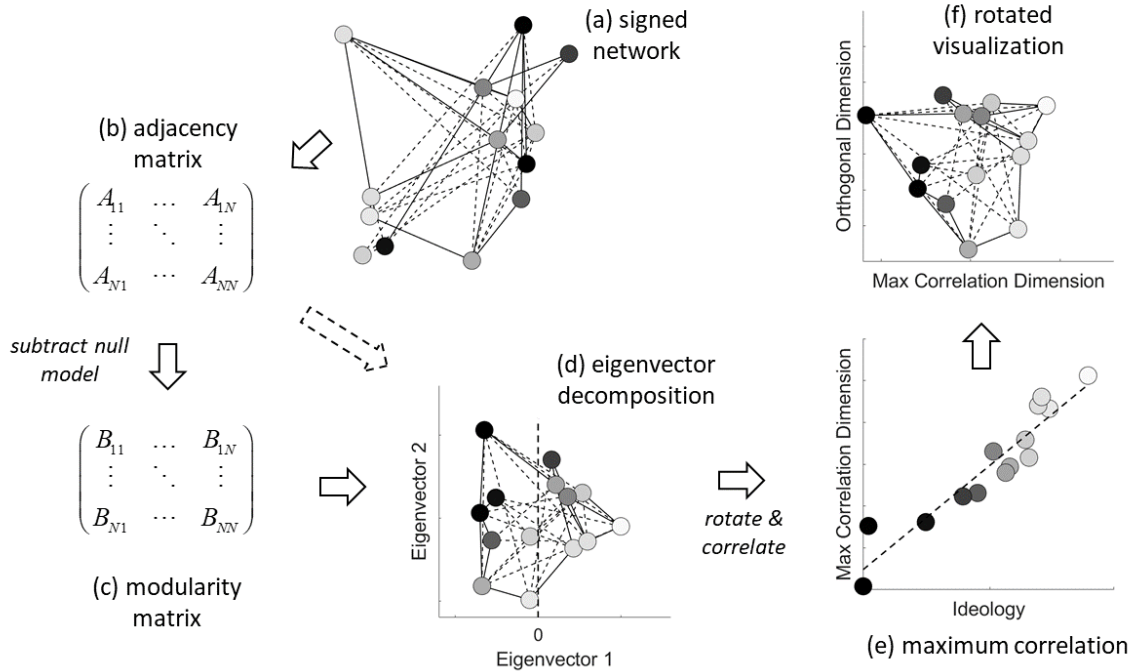


Figure 4. Overview of signed network analysis procedure involving eigenvector decomposition of adjacency or modularity matrices. Solid and dashed lines correspond to positive and negative ties respectively. Lighter nodes have higher ideology values. See Ref. [5] for details.

3.4 Drivers of Militant Cooperation and Conflict

3.4.1 Syria

Using the network analysis methods described above, we investigated the importance of ideology, power, and state sponsorship as drivers of cooperation and conflict in Syrian militant networks [Refs. 1 and 2]. Ideology was found to be strongly related to network structure. We conceptualized ideology into three distinct dimensions – conflict frame, ideal polity, and territorial aspiration – and tested them separately as well their average. For the core cooperation network (consisting of 31 groups with at least one joint operation), we found that groups who were close on conflict frame, territorial aspiration, and average ideology were more likely to cooperate. For the core infighting network (consisting of 30 groups with at least one episode of infighting), ideologically distant groups were more likely to fight. This was the case for average ideology as well as each dimension separately.

With respect to power, for both cooperation and infighting, we tested opposed hypotheses of power symmetry and power asymmetry, involving the interaction of groups with similar and disparate sizes respectively. For cooperation, we found ambiguous support for a power symmetry effect in which strong groups cooperate more with other strong groups weak with other weak groups. However, for infighting, the hypothesis of power asymmetry (strong groups fighting weak groups) was supported by the assortativity, latent space models, and ERGMs. We also tested the hypotheses that having a common state sponsor would encourage cooperation and discourage infighting. No support for either state sponsorship effect was found.

While the analyses published in Refs. [1] and [2] treated cooperation and conflict separately, it is also possible to combine the networks into a single signed network as shown in Figure 5. The first eigenvector, along the vertical axis, represents the inter-rebel war that took place starting in 2014 in which ISIL fought with many other insurgent groups. The second eigenvector reflects the operation of overall ideological homophily, with horizontally proximate groups more likely to cooperate and less likely to fight.

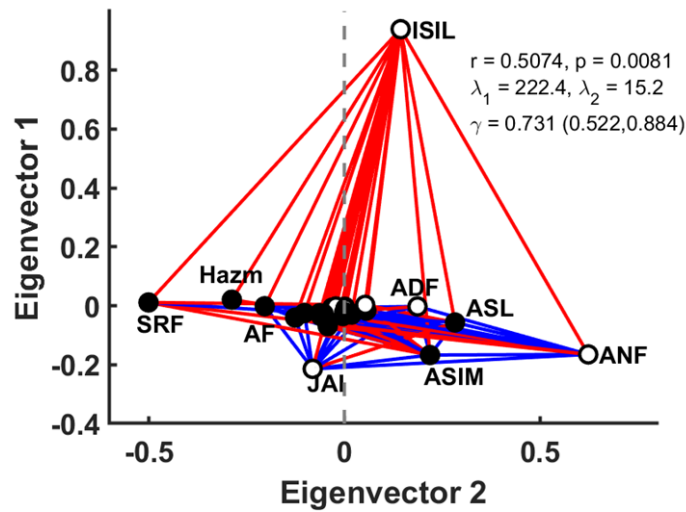


Figure 5. Signed network of cooperation and conflict among 31 Syrian militant groups from 2013-2015. Cooperative ties in blue, conflictual ties in red. Nationalist groups in black, jihadists in white. r is the correlation of average ideology with group coordinates along Eigenvector 2 (p is p -value). λ_1 and λ_2 are the first and second eigenvalues.

3.4.2 Ukraine

We used the network constructed from the government-translated documents to further develop signed network methods and investigate the drivers of militant factional structure [Ref. 5]. We developed a theory as to whether and how ideology structures patterns of insurgent conflict and cooperation when state sponsorship is crucial for group survival. The theory encompasses the multiple and single state sponsor cases. The case of a single state sponsor is the situation in which ideology would be thought least likely to matter. In this case, logic suggests that ideology should not shape the militant network as the only viable militant groups should be those whose stated ideology matches the positions of the external sponsor, therefore garnering material support and the advantages it confers such as weapons, funds, training, and safe havens. However, we theorize that it is still possible to observe a range of ideological positions in the single state sponsor case. This can occur due to the multidimensional nature of militant ideology. If there are dimensions of ideology along which the external state sponsor is ambiguous, militant groups can differentiate themselves along those dimensions while still receiving external material support. Those ideological differences can then shape patterns of intra-movement cooperation and conflict in accord with the operation of homophily.

In Ukraine, Russia was the single state sponsor of the separatists and its support was crucial to the survival of the insurgency. Consequently, this case represents a hard test of the ability for ideology to structure militant networks given that initial power-centric intuition would hold that ideology would be unimportant for militant factionalism. Our signed network analysis employed the sign permutation test of assortativity described above as well as both modularity-based and adjacency matrix-based spectral representations of community structure. The analysis shows that one dimension of the three in our militant ideology framework – territorial aspiration, operationalized as support for forming the breakaway region of “Novorossiya” – does structure the network in accordance with ideological homophily. In line with our theory, it was regarding this dimension

that the state sponsor, Russia, sent ambiguous messages concerning its preferences. In contrast, Russia had a clear position on the conflict frame, Russia vs. the West, which did not exhibit homophily. Figure 6 shows how the network of militant cooperation and conflict during the first phase of the insurgency, in which Russia's position on secession was ambiguous, is structured by the level of support for Novorossiia.

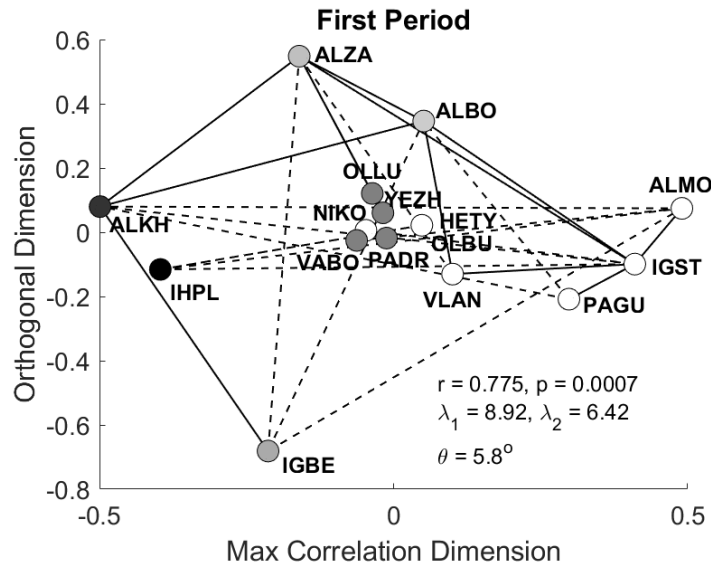


Figure 6. Structure of network of cooperation and conflict among pro-Russian militant leaders in Ukraine from Feb. 2014 to May 2015. Positive and negative ties shown as solid and dashed lines respectively. Lighter shades indicate more support for seceding to form Novorossiia. r is the correlation of support for Novorossiia with node coordinates along the axis of maximum correlation (p is p-value). θ is the angle of rotation. λ_1 and λ_2 are the first and second eigenvalues.

3.5 Modeling Signed Network Structure and Dynamics

3.5.1 Phase Transitions in Signed Networks

Our empirical representation of militant networks uses signed networks in which cooperative interactions are positive ties and conflictual interactions are negative ties. As modeling methods for signed networks are relatively undeveloped compared to unsigned networks, we investigated some fundamental aspects of signed network behavior. One aspect concerned community structure, which in the militant context could correspond to different ethnic, religious, or ideological identities. Using a stochastic simulation to generate signed networks with community structure, we discovered two types of significant phase transitions in network structure as reflected in their eigenvalue spectrum. One type is the “detectability transition,” known to occur in unsigned networks, which involves a sudden transition in the ability to detect community structure. The other type is the “sociality transition,” which is unique to signed networks and involves the ability to discern whether an overall tendency toward positive (prosocial) or negative (antisocial) tie formation is present. Both types of transitions occur as parameters controlling tie formation probabilities are varied. We derived mathematical expressions

accurately predicting where these transitions occur using perturbation analysis and random matrix theory. This work was published in Ref. [4].

3.5.2 Link between Phase Transitions and Structural Balance Evolution

The detectability and sociality transitions discovered above were found to induce transitions in the evolution of the network under structural balance dynamics. Structural balance theory, the most prominent theory of signed social networks, states that only two types of triads are stable: three positive ties (the friend of my friend is my friend) and two negative/one positive (the enemy of my enemy is my friend). We used the stochastically-generated networks with community structure as initial conditions for a simple differential equation model of structural balance dynamics. This model has only two types of outcomes: (1) a factionalized state consisting of two mutually hostile factions, each with only positive intra-faction ties and only negative inter-faction ties; and (2) a harmonious state in which all nodes are positively connected. The detectability transition governs the dynamical transition in the composition of the factions in the factionalized state: for sufficiently strong initial community structure, the factions completely correspond to the initial identity blocks so that the network becomes polarized by identity, whereas for weaker structure the factions are of mixed identity. The identity polarized state points to a lock-in effect in which initial identity divisions are amplified as has been observed in ethnic conflicts. The sociality transition governs the transition between the mixed identity factionalized state and the harmonious state. Figure 7 displays examples of how different initial network structure leads to the three different final states and how the phase transition behavior above leads to sharp boundaries between these regimes in parameter space. This work appeared in Ref. [4].

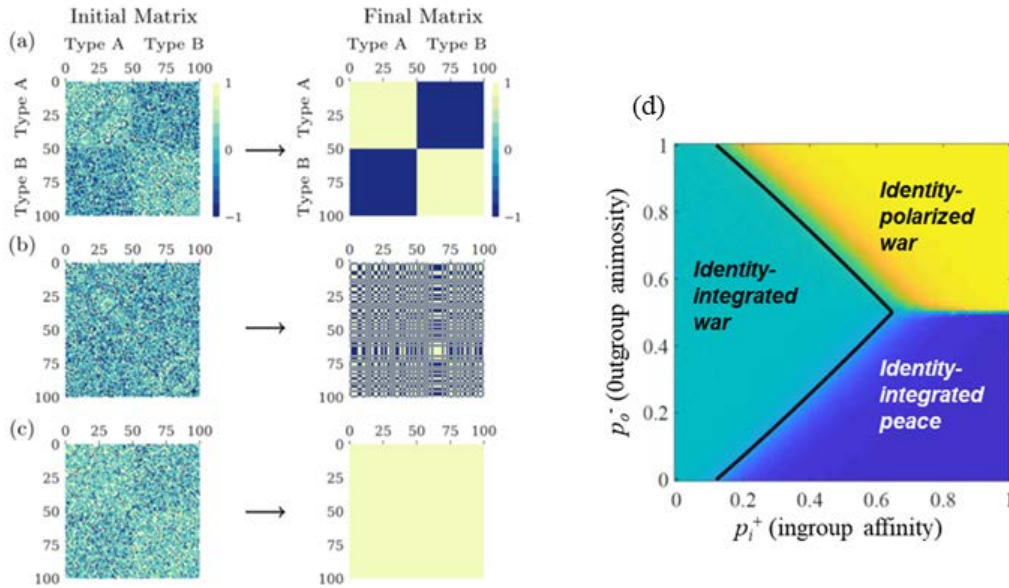


Figure 7. Evolution of networks under structural balance dynamics. (a) An initial network consisting of strong community structure corresponding to two different identity types evolves into a final state polarized by identity. (b) A network with weak initial identity structure evolves into a two-faction state where each faction consists of mixed identities. (c) A network with a strong overall tendency to form positive ties evolves into a harmonious single-faction of mixed identities. (d) Phase diagram showing sharp boundaries separating the different final state regimes. The ingroup and outgroup affinities control the probabilities of forming positive ties among nodes of the same and different identity blocks respectively. The solid black lines are the theoretical predictions for the boundaries.

3.6 Tracking and Anticipating Militant Conflict

3.6.1 Structural Balance

Using our Syria data, we investigated the extent to which militant networks can be said to obey structural balance theory. We used a recently developed method that allows for evaluation of whether structural balance is statistically significant (Kirkley, Cantwell and Newman, Physical Review E, 2019). We found that the network aggregated over the whole time period of the data was significantly balanced both in the weak sense (all negative triad allowed as balanced) and the strong sense (all negative triad not allowed). We also disaggregated the network by six-month time intervals and found that all of the time intervals were significantly balanced (see Figure 8). This is the first finding of structural balance in militant networks.

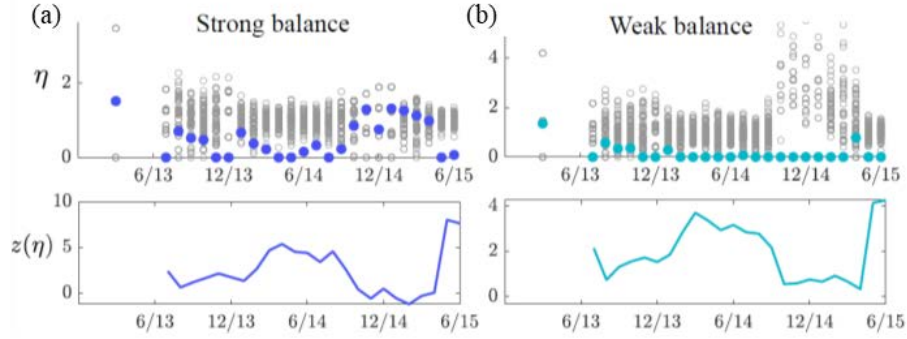


Figure 8. Structural balance level of Syrian militant network over time. (a) Strong balance. (b) Weak balance. Top panels compare observed network balance metric (blue, cyan stars) against metrics from null simulation runs (gray circles). Lower η values indicate the network is more balanced. Bottom panels combine observed and simulated η values into z-scores where higher values indicate more balance.

3.6.2 Spectral Analysis and Eigenvector Polarization

Motivated by our research on fundamental signed network behavior above, we investigated the spectrum of our Syria data. We focused on the behavior of eigenvectors of the adjacency matrix residing outside the main spectral “noise” band (Figure 9a). These eigenvectors were found to track important aspects of the system including insurgent group ideology and power distribution. The magnitude of the first adjacency eigenvalue was found to correlate with the level of structural balance in the system.

Based on the relationship of the first adjacency matrix eigenvalue to the level of structural balance, we developed a novel “eigenvector polarization” metric, which can be used to track the level of conflict in the network and detect the onset of systemic wars. Structural balance is not a good metric of conflict in signed networks because it can be high for opposing situations of global harmony and factional polarization. The same holds true for the first eigenvalue, which can be large and outside the noise band for both global peace and conflict. However, the components of the first eigenvector are different in the two situations: the components carry community structure information when the network is polarized but not when it is in the global harmony state. The amount of community structure contained by the first eigenvector, or any eigenvector, can be gauged by finding its matrix element with respect to the modularity matrix $\mathbf{M}$. We therefore define the eigenvector polarization, φ_i , of the i^{th} eigenvector, $\mathbf{v}_i$, by

$$\varphi_i = \mathbf{v}_i^T \mathbf{M} \mathbf{v}_i,$$

where T indicates transpose. When the first eigenvector polarization, φ_1 , is close to zero, the network is in a harmonious state. When it is high, the network is polarized.

Application to the Syrian militant network shows the behavior of the adjacency matrix spectrum, modularity matrix spectrum, and eigenvector polarizations over time (Figure 9). The first adjacency eigenvalue is far outside the noise band edge, which is estimated using random matrix

theory, early in the time period in Figure 9a, indicating a high level of structural balance. The first modularity eigenvalue in this same period, as seen in Figure 9b, is not clearly distinct from the estimated noise band edge implying relatively weak community structure. The first eigenvector polarization, ϕ_1 , is seen to be close to zero during this time period (Figure 9c), indicating that the high level of structural balance corresponds to a harmonious state and not a polarized one. This is also evidenced by the first eigenvector components (not shown), which are of uniform sign and thereby carry no community structure. This early period was marked by good overall cooperation among militant groups.

In contrast, the period after January 2014 (point A in the figure) has both first adjacency and modularity eigenvalues well outside the noise band. The first eigenvector polarization is also outside the noise band as can be seen in Figure 9c. This period thereby corresponds to systemic conflict among the militant groups. Given that its value relative to the noise band edge is of key significance, we normalize the eigenvector polarization by the band edge as plotted in Figure 9d. The normalized first eigenvector polarization first exceeds a value of one in January 2014, which is when a collection of rebel groups launched an offensive directed at ISIL, thereby initiating an inter-rebel war. This represents an empirical confirmation of the ability of the eigenvector polarization to detect the onset of systemic conflict. Follow-on work (performed under a grant from the Army Research Office) using the network of European international relations further verified this capability in that the eigenvector polarization was better able to detect conflicts between great powers than metrics based on the adjacency matrix, modularity matrix, or structural balance alone.

A paper describing the structural balance and eigenvector polarization results is in preparation.

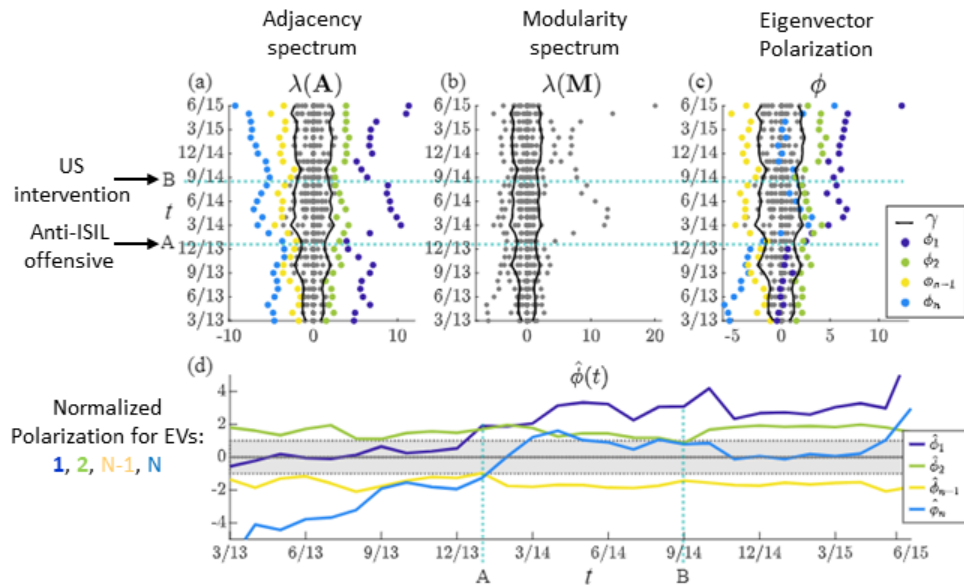


Figure 9. Spectral analysis of Syrian militant cooperation and conflict network over time. (a) Signed adjacency matrix eigenvalues. Blue dots correspond to first eigenvector, green is second, yellow is second to last, cyan is last. Estimated edges of main spectral band obtained from random matrix theory indicated by solid black lines. (b) Modularity matrix eigenvalues. (c) Eigenvector polarizations. (d) Normalized eigenvector polarization metrics over time for two highest and two lowest eigenvectors. First eigenvector polarization (blue) is close to zero at the beginning of the time period indicating largely harmonious inter-militant relations but it crosses the noise band threshold (top edge of gray band) coincident with the start of the inter-militant war (marked by A).

3.6.3 Predicting Conflict Links

We investigated the ability to predict future instances of conflict between militant groups given past conflict ties using different machine learning approaches following earlier work done on interstate conflict (Cranmer and Desmarais, Political Analysis, 2017). We found good performance for the Syria data, which was substantially better than the ability to predict interstate conflict. We have also evaluated these methods on data from the Libyan civil war collected from the ACLED database and good performance was obtained although less than the Syrian case. Application to the Ukraine data, however, showed even worse performance than the interstate case.

4 Transition Activity

The following is a list of the activity related to potential transition of this research:

1. In July 2017, PI briefed Defense Threat Reduction Agency (DTRA) personnel on this research.
2. In December 2018, PI briefed US government analysts and methodologists on this research.

3. In December 2019, PI briefed US government analysts and methodologists on this research.
4. In November 2020, PI gave an online presentation about this research to US government analysts and methodologists.

5 Training

The following University of Washington graduate students received support as research assistants under this grant: Emily Kalah Gade (Political Science), Calvin Garner (Political Science), Alexis McClimans (Information School), Megan Morrison (Applied Mathematics), Nora Webb Williams (Political Science).

The following UW undergraduate students received support as research assistants: Matthew De La Roca and Elizabeth McKinnon.

6 Publications and Presentations

6.1 Publications

1. E.K. Gade, M.M. Hafez, and M. Gabbay, “Fratricide in rebel movements: A network analysis of Syrian militant infighting,” *Journal of Peace Research*, 56(3), 321-335, 2019 doi:10.1177/0022343318806940 (Appendix 1)
2. E.K. Gade, M. Gabbay, M. M. Hafez and Z. Kelly, “Networks of Cooperation: Rebel Alliances in Fragmented Civil Wars,” *Journal of Conflict Resolution*, 63(9), 2071-2097, 2019. doi:10.1177/0022002719826234 (Appendix 2)
3. M. Gabbay, “Integrating Computational Modeling and Experiments: Toward a More Unified Theory of Social Influence,” in *Social-Behavioral Modeling for Complex Systems*, P. Davis, A. O’Mahony, J. Pfautz (eds.), Wiley, 2019. doi:10.1002/9781119485001.ch1 (Appendix 3)
4. M. Morrison and M. Gabbay, “Community detectability and structural balance dynamics in signed networks,” *Physical Review E*, 102(1), 012304, 2020. doi:10.1103/PhysRevE.102.012304 (Appendix 4)
5. N. Webb Williams, C. Garner, and M. Gabbay, “Militants and Mixed Messages: The Effect of Sponsor Ambiguity on Insurgent Ideology and Networks,” preprint, 2021. Available at SSRN: <https://ssrn.com/abstract=3791678> (Appendix 5)

6.2 Presentations

Presenter(s) listed in boldface.

1. **Gade, E.K., Gabbay, M.**, Hafez, M., and Kelly, Z., “Bringing Ideology Back In: Militant Networks and Alliance Formation in Syria’s Civil War,” Workshop on Ideology and Armed Groups, February 2017, Baltimore, MD.
2. **Gade, E.K.**, Gabbay, M., **Hafez, M.**, and Wilkerson, C., “Infighting, Ideology, and Network Structure in the Syrian Insurgency,” International Studies Association Annual Meeting, February 2017, Baltimore, MD.
3. **Gade, E.K.**, Hafez, M., and Gabbay, M., “My Brother’s Reaper: The Targets of Fratricide in Insurgent Networks,” 2nd Workshop on Ideology and Armed Groups, June 2017, Montreal.
4. **Gabbay, M.**, Gade, E.K, Hafez, M., and Kelly, Z., “Cooperation, Infighting, and Network Structure in the Syrian Insurgency,” North American Social Networks Conference, July 2017, Washington, DC.
5. **Gabbay, M.**, “Extremism on Networks and Networks of Extremism,” Defense Threat Reduction Agency, July 2017, Fort Belvoir, VA.
6. **Gade, E.K., Hafez, M.**, and Gabbay, M., “Fratricide in Insurgency,” American Political Science Association Annual Meeting, August 2017, San Francisco, CA.
7. **Gade, E.K., Hafez, M.**, and Gabbay, M., “Fratricide in Rebel Movements: A Network Analysis of Syrian Militant Infighting,” International Studies Association Annual Meeting, April 2018, San Francisco, CA.
8. Williams, N. W. and **Gabbay, M.**, “Militant Dynamics among Eastern Ukrainian Separatists,” International Studies Association Annual Meeting, April 2018, San Francisco, CA.
9. **Gabbay, M.**, “Modeling Militant Factional Dynamics,” North American Social Networks Conference, November 2018, Washington, DC.
10. **Williams, N. W.**, Garner, C., and Gabbay, M., “Cooperation, Conflict, and Ideology within Militant Networks in Eastern Ukraine,” International Studies Association Annual Meeting, March 2019, Toronto, Canada.
11. **Morrison, M.** & Gabbay, M., “Community Structure and Structural Balance in Signed Networks,” poster presented at SIAM Workshop on Network Science, May 2019, Snowbird, UT.
12. **Gabbay, M.**, “A Nonlinear Model of Opinion Network Dynamics and Its Experimental Investigation,” SIAM Conference on Applications of Dynamical Systems, May 2019, Snowbird, UT.
13. **Morrison, M.** & Gabbay, M., “Structural Balance Dynamics in Insurgent Networks,” SIAM Workshop on Network Science, July 2020 (online).
14. Webb Williams, N., **Garner, C.**, & Gabbay, M., “Cooperation, Conflict, and Ideology within Militant Networks in Eastern Ukraine,” American Political Science Association Annual Meeting, September 2020 (online).
15. Morrison, M. & **Gabbay, M.**, “A Spectral Measure of Polarization in Conflict Networks,” Networks 2021, July 2021 (online).

16. Morrison, M. & **Gabbay, M.**, “A Network Measure of Systemic Conflict in International Relations and Insurgency,” American Political Science Association Annual Meeting, October 2021, Seattle, WA.

Appendix 1

Fratricide in Rebel Movements: A Network Analysis of Syrian Militant Infighting

Fratricide in rebel movements: A network analysis of Syrian militant infighting

Emily Kalah Gade

Department of Political Science, University of Washington

Mohammed M Hafez

Department of National Security Affairs, Naval Postgraduate School

Michael Gabbay 

Applied Physics Laboratory, University of Washington

Journal of Peace Research
2019, Vol. 56(3) 321–335
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0022343318806940
journals.sagepub.com/home/jpr



Abstract

Violent conflict among rebels is a common feature of civil wars and insurgencies. Yet, not all rebel groups are equally prone to such infighting. While previous research has focused on the systemic causes of violent conflict within rebel movements, this article explores the factors that affect the risk of conflict between pairs of rebel groups. We generate hypotheses concerning how differences in power, ideology, and state sponsors between rebel groups impact their propensity to clash and test them using data from the Syrian civil war. The data, drawn from hundreds of infighting claims made by rebel groups on social media, are used to construct a network of conflictual ties among 30 rebel groups. The relationship between the observed network structure and the independent variables is evaluated using network analysis metrics and methods including assortativity, community structure, simulation, and latent space modeling. We find strong evidence that ideologically distant groups have a higher propensity for infighting than ideologically proximate ones. We also find support for power asymmetry, meaning that pairs of groups of disparate size are at greater risk of infighting than pairs of equal strength. No support was found for the proposition that sharing state sponsors mitigates rebels' propensity for infighting. Our results provide an important corrective to prevailing theory, which discounts the role of ideology in militant factional dynamics within fragmented conflicts.

Keywords

civil war, fragmentation, ideology, infighting, social network analysis, Syria

A revolutionary's worst enemy is often another revolutionary.

Lichbach (1995: 203)

infighting occurs frequently. The history of civil conflicts is replete with dramatic instances of rebel-on-rebel fratricide (Bakke, Cunningham & Seymour, 2012).² The

Introduction

Infighting among rebels is a common feature of civil wars and insurgencies. Rebel movements are usually divided into brigades that fight under several factional banners with varying degrees of coordination.¹ This fragmentation generates a competitive landscape in which violent

² Iconic episodes of interrebel fratricide include Stalinists against Trotskyists during the Spanish Civil War (May 1937); Yugoslavian Communist Partisans against the Nationalist Chetniks during World War II; the Algerian National Movement against the National Liberation Front during their war of independence from France (1954–62); and Al-Qaeda against Iraqi Islamists and nationalists during the US occupation of Iraq (2003–11).

¹ Of 181 insurgencies since 1946, more than half involved multiple insurgent groups. Since the 1980s, 64% involved multiple rebel factions (Jones, 2017: 168).

Corresponding author:

ekgade@uw.edu

ongoing Syrian civil war offers a stark reminder of how rebels can turn on each other while simultaneously waging war against a formidable regime.

The puzzle of rebel infighting can be addressed at either the systemic or dyadic level. The burgeoning literature on interrebel wars almost exclusively focuses on systemic risks that generate conflictual rebel relationships. These include the problem of credible commitments born out of anarchy (Christia, 2012), the depth of movement fragmentation (Cunningham, Bakke & Seymour, 2012), regime weakness or impending rebel victory (Lichbach, 1995), the presence of lootable resources (Fjelde & Nilsson, 2012), and the quest for patronage within violent patrimonial political systems (Seymour, 2014). In this article, we take a dyadic approach to understanding which groups are most prone to infighting when rebel movements descend into factional conflicts. Investigation at the dyadic level helps us go beyond the systemic assumption of unit homogeneity and thus can offer finer predictions about who is likely to enter into fratricidal wars.

We make three contributions – theoretical, methodological, and empirical. Theoretically, we investigate the effects of *power*, *ideology*, and *state sponsorship* on the propensity for infighting in rebel dyads. We conceptualize power in conflict dyads as either *symmetric* or *asymmetric* (i.e. groups of similar or dissimilar strength, respectively). Whereas power parity may generate conflict between an established rebel faction and a rising competitor, power asymmetry may invite rebel aggression by strong factions against their weaker rivals. We test both propositions. We conceive of ideology in conflict dyads as either *proximate* or *distant* (i.e. groups with overlapping ideological positions or opposing ideological preferences, respectively). We hypothesize that higher rates of conflict are more likely between ideologically distant groups than between those that are ideologically proximate. Lastly, we explore the potential effects of state sponsorship on interrebel conflicts by looking for the presence or absence of *overlapping state sponsors* in rebel dyads. We posit that rebels that share state sponsors are incentivized by their external patrons to forge unity and will thus experience less infighting than those with distinct state sponsors.

Methodologically, we introduce a network-analytic approach to explain the determinants of salient conflict dyads in rebel movements. Conflict dyads can be embedded within a network of movement infighting, which enables the detection of patterns of infighting relationships. We use assortativity, community structure detection, and network simulation, as well as an additive

and mixed effects (AME) latent factor model to evaluate the effects of power, ideology, and state sponsorship on generating conflict within dyads. We also run a number of robustness tests including validation of our findings using exponential random graph models (ERGM), another approach to statistical inference on networks (Desmarais & Cranmer, 2017).

Empirically, we rely on relational and quantitative analysis of the ongoing Syrian civil war, a conflict that is at the center of regional and international insecurity, multiple humanitarian crises, and military interventions by major powers. We constructed a unique database of three years of rebel infighting for the period of January 2013 to December 2015. The data come from primary insurgent documents such as rebel operational communiques, social media postings, and jihadist web forums. We also drew upon rebel groups' political programs and manifestos to capture their ideological leanings. Lastly, we collected data on power measures, operational location, and state sponsors of the most prominent groups involved in interrebel conflicts. This rich dataset allows us to test our hypotheses with a number of robustness checks to bolster the validity of our conclusions.

We find compelling evidence that ideology is a major driver of infighting in rebel movements. An ideological difference among rebel dyads consistently increases their propensity for infighting, a result that is robust across analyses and time. Specifically, sectarian jihadists were the most prone to engage in interrebel wars and they did so mainly with non-sectarian Islamists and with secular nationalists and Kurdish separatists. Comparatively, groups that were nationalists or Kurdish separatists did not fight among each other as often as groups of different ideological types. We also observe strong, although less consistent, evidence for power asymmetry infighting dyads; Syrian rebel infighting is usually between groups of disparate strengths. Lastly, we found no relationship in any model between state sponsorship and infighting. Despite the rivalry among the rebel sponsoring states, we could not find clear evidence that this rivalry shaped militant infighting patterns in the Syrian conflict.

Power in conflict dyads

In an anarchic context with no central authority to enforce binding promises within rebel movements, information and credible commitments problems force rebel groups to be self-regarding and consider their survival above all else (Christia, 2012). Relative power considerations can determine who falls victim to aggression, who merely survives, and who ultimately thrives and

captures the lion's share of post-conflict spoils (Krause, 2017). Relative power considerations can lead to two predictions about which rebel dyads are more likely to fight. When rebels confront each other, their power distribution can be either *asymmetric* (one group is substantially more powerful than the other) or *symmetric* (both groups are roughly equal in capabilities).³ Both scenarios are capable of generating interrebel conflicts.

Powerful groups can exploit the asymmetry in forces by eliminating minor players that infringe on their territory and resources (Fjelde & Nilsson, 2012). They may also attack weaker groups that hold the potential to grow in power and thus challenge their leadership in the future (Pischedda, 2018). Strong rebel groups can also target weaker factions that may act as spoilers in conflict-ending negotiations. Although less intuitive, it is possible for weaker actors to undertake the risk of challenging powerful rebel groups because the payoff is quite high if they are successful (Krause, 2017). This 'gamble for resurrection' is especially likely if the minor challenger is going after a powerful group that controls a resource-rich territory, which can rapidly accelerate the ascendancy of the weaker party (Fjelde & Nilsson, 2012). Thus, our first power hypothesis:

Hypothesis 1a: Asymmetric power – Infighting will be more likely between groups of disparate power.

In contrast, symmetric power distribution can produce infighting among rebels as an emerging and dissatisfied militant organization approaches parity with an established rebel group. Two equally powerful rebel groups could threaten one another's security and leadership aspirations, so power parity is a cause for concern for established rebel organizations (Krause, 2017). Two mechanisms help explain how power parity can unleash interrebel violence. The disruption to the existing rebel power hierarchy leads to greater conflict as the hegemonic rebel group, feeling threatened by a rapidly rising rebel faction, seeks to prevent the latter's continued ascendancy. Or, the newly ascendant rebel power itself could initiate conflict by challenging the status quo under the hegemonic rebel faction because it seeks greater representation

within the rebel institutional hierarchy (McLauchlin & Pearlman, 2012). Thus, our second power hypothesis:

Hypothesis 1b: Symmetric power – Infighting will be more likely between groups of comparable power.

Ideology in conflict dyads

Rebel groups are fragmented along their ideological preferences, not just their power capabilities. Ideology reflects a group's political demands, normative commitments, and future objectives. It also helps bind rebels to their commanders by fostering identification with group goals and it can motivate commitment and sacrifice (Lichbach, 1995: 92–93). That is why insurgent organizations from diverse traditions – Marxists, Maoists, ethnonationalists, and fundamentalists – dedicate time and resources to socialize their recruits ideologically (Oppenheim et al., 2015; Hoover Green, 2016). We would expect that under scope conditions of ideological diversity, competition and conflict will shape interrebel relationships (Seymour, Bakke & Cunningham, 2016).

Following Gutiérrez Sanín & Wood (2014: 215), we define ideology in rebel movements as:

a systematic set of ideas that includes the identification of a referent group (a class, ethnic, or other social group), an enunciation of the grievances or challenges that the group confronts, the identification of objectives on behalf of that group (political change – or defense against its threat), and a [...] program of action.

We operationalize this definition along three dimensions: *conflict framing*, *ideal polity*, and *territorial aspiration*. Conflict framing specifies the primary referent group for which rebels are fighting, and the out-groups they find most threatening. This is particularly important for conflicts with multiple identity groups, which is common in multi-ethnic civil wars. Ideal polity refers to the nature of the post-conflict political order that rebel groups aim to create. This dimension captures traditional right-left ideological divides as well as divisions between those seeking to create secular or fundamentalist polities. Territorial aspiration refers to the boundaries of the ideal polity, addressing the core debate between those who wish to maintain the territorial integrity of their states and those who seek to break up the polity into multiple states. Movements with shared conceptions of the ideal polity sometimes diverge over the territorial boundaries

³ When referring to power, we mean relative power capabilities as measured by estimates of the group size. Although an imperfect measure, group size is often used in large-N statistical analyses (Akcinaroglu, 2012; Christia, 2012; Krause, 2013/4). We make the assumption that group size is a proxy for other elements of rebel power, such as financial resources.

of that polity.⁴ Territorial aspirations have been at the root of many secessionist civil conflicts, resulting in 131 sovereign states coming into existence since 1945 (Griffiths, 2016).

We hypothesize that divergence along these three ideological components can aggravate infighting in rebel dyads. Conversely, group dyads with similar ideological positions along these three dimensions will exhibit low rates of infighting. Thus, our ideological hypothesis:

Hypothesis 2: Ideological distance – The greater the ideological distance between two rebel groups, the higher the likelihood of infighting.

Three causal mechanisms help explain how ideological differences can produce infighting. First, groups with fundamentally divergent conceptions of the ideal polity are likely to view their cohabitation in the rebel field as mutually threatening. Not only do their competing ideological visions violate their core normative commitments for which they are making the ultimate sacrifice, their divergent conflict objectives make credible commitments difficult to uphold. Absent trust, competing camps see their coexistence as a zero-sum game with little possibility for power-sharing in the future.

In contrast, groups with shared conceptions of the ideal polity corroborate each other's core political preferences and thus can readily signal to their ideological kin their intentions to share power in the post-conflict political order. Moreover, the ascendancy of ideologically similar groups is less threatening to one's core constituency and sponsors, reducing the pressure to compete over leadership. Conversely, a conflict between ideologically similar groups can expose the conflicting parties to condemnation from their supporters because their infighting undermines the unity of their ideological faction.

A second ideological mechanism involves the relationship between conflict framing and targeting. Groups with opposing conflict narratives are likely to adopt divergent targeting portfolios (Gutiérrez Sanín & Wood, 2014). For instance, an overtly sectarian conflict frame may justify expansive attacks against the civilians of a rival sect more readily than a frame that rejects sectarian divisions and, instead, paints all of the nation and its diverse sects as equally vested in forging a new polity.

⁴ For example, Islamists today are divided between those who favor establishing an Islamic order within the modern national state and those that harbor the irredentist ambition of restoring an Islamic caliphate.

The debate over the legitimate targets of violence is often a key source of dissension within rebel movements, inviting open conflict.

A third ideological mechanism relates to competing visions of territorial sovereignty. As a conflict becomes protracted, the territorial integrity of the state may become a subject for negotiation. Rebels that harbor broader or narrower territorial ambitions may clash with rivals that seek to maintain the extant state boundary. Removing spoilers from the rebel movement can thus drive interrebel conflicts.⁵

State sponsorship in conflict dyads

State sponsorship can generate both rebel unity and rebel rivalries. Civil wars invariably invite external actors to intervene on behalf of the combatants, seeking to project influence and prevent rival states from adversely shaping the conditions for conflict termination.⁶ The sponsorship of proxy actors is a cost-effective way for states to compete with their state rivals (Salehyan, 2010). External patrons thus provide arms, money, supplies, or sanctuaries to rebel groups in the expectation that these rebels will exhibit sufficient discipline and cohesion to fulfill their patron's strategic aims. Sponsors can threaten to withhold financing and war materiel from those who are jeopardizing a cohesive rebel coalition (Lichbach, 1995: 179). Bapat & Bond (2012) view such external leverage as an important interrebel institution that can help overcome the credible commitments problem, increase cooperation, police against side negotiations, and mediate conflicts between rebel groups.

However, state sponsors can also undermine rebel unity by incentivizing some rebels to challenge their rivals (Tamm, 2016). This is particularly the case when multiple state sponsors with opposing political agendas seek to foster their own proxy clients through patronage. The presence of multiple sponsors increases the degrees of freedom rebel groups can exercise to support themselves and reduces the degrees of freedom any individual external patron can exert to foster cohesive rebel coalitions (Salehyan, Siroky & Wood, 2014). Thus, whereas overlapping state sponsorship in rebel dyads should

⁵ For example, the conflict between Hamas and Fatah during the 1990s revolved around the former's refusal to accept a two-state solution. Hamas sought to sabotage the peace process through suicide attacks, which led the Fatah-led Palestinian Authority to crack down on Hamas (Kydd & Walter, 2002).

⁶ According to Jones (2017: 136), of 181 insurgencies between 1946 and 2015, 82% involved outside support.

mitigate conflict, this moderating effect is diminished in dyads with non-overlapping state sponsors.

Hypothesis 3: Overlapping state sponsorship – Rebel groups that derive support from the same state sponsors will experience less infighting than those who have distinct state benefactors.

Network analytic methodology

Social network analysis in political science is designed to account for interdependence within a system of political actors (Ward, Stovel & Sacks, 2011; Hafner-Burton, Kahler & Montgomery, 2009; Maoz, 2010). Dyadic models that assume independence of observations mischaracterize relational data because infighting is dependent upon relationships with a range of groups within the system (Dorussen, Gartzke & Westerwinter, 2016). For example, network analysis can account for how variation among groups in their degree of infighting (their total number of clashes with other groups) constrains the overall pattern of infighting. The relative numbers of high and low degree groups shape the extent to which high degree groups must primarily fight each other or can fight with many lower degree groups instead. The fragmented nature of asymmetric conflicts makes network analysis a promising quantitative approach for evaluating militant behaviors in multiparty wars (Zech & Gabbay, 2016).

A social network consists of nodes and the ties between pairs of nodes. The nodes can be individuals, organizations, or countries and the ties signify relationships such as communication, cooperation, or conflict. In our empirical analysis, we employ a network in which ties represent the number of infighting episodes between group dyads. Four different methods of investigating and testing the relationship between the conflictual network and the independent variables are applied: (1) comparing the ‘assortativity’ – a measure of the variable’s tendency toward producing homophily or heterophily (connection of like or unlike nodes, respectively) – of the observed network with the assortativity distribution obtained from a null model simulation; (2) correlating the variable with important patterns in the network as found via eigenvector-based representations of community structure; (3) a simulation of tie formation that explicitly includes the variable to estimate the characteristic zone within which conflict is enhanced (homophily) or suppressed (heterophily); and (4) an additive and multiplicative effects (AME) latent factor model, relying on the Markov Chain Monte Carlo algorithm. The first three

are used to analyze power and ideology separately and the fourth allows us to compare power, ideology, and state sponsorship simultaneously.

Assortativity is the standard measure used to assess whether tie formation is driven by similarity with respect to a scalar variable (as we operationalize power and ideology). The assortativity is the correlation of the variable values at each end of a tie taken over all ties (see Appendix). An assortativity value of +1 corresponds to a network with maximal homophily whereas –1 signifies maximal heterophily. For statistical testing purposes, the assortativity cannot be treated as one would treat a standard correlation because ties are not taken to be independent. Accordingly, we compare the observed assortativity with the distribution obtained from a null model simulation in which the independent variable of interest is not included: if the observed value is greater (less) than the simulation mean then the tie formation process exhibits homophily (heterophily).

Network structure can be visualized in a way that relates to the assortativity of the variable of interest. The modularity matrix is a transformation of the tie data (see appendix) that is often used for community detection purposes (Newman, 2006). Its eigenvector decomposition can be used to identify patterns of tie formation that are shaped by the variable. If tie formation displays homophily with respect to the variable, then the variable should correlate to some extent with one of the highly ranked (most positive eigenvalues) eigenvectors. However, if tie formation displays heterophily, the variable should instead correlate with one of the lowest-ranked (most negative eigenvalues) eigenvectors (Newman, 2006).

In the null simulation, nodes form ties (fight with each other) probabilistically. Each iteration consists of the placement of a tie between nodes where the iterations proceed up to the total number of ties in the observed network. The simulation seeks to reproduce the observed node degrees and so assumes that the propensity of a group to fight with other groups is known but not the distribution of its infighting ties. As a result, each node can only receive a maximum number of ties equal to its observed degree (it is not always possible to reproduce the degrees exactly, but the differences are typically small). At any given iteration, the degree deficit by which a node’s current degree falls short of that maximum affects its tie formation probability – the larger the degree deficit, the more likely it will form a tie. The degree deficit decreases until it reaches zero, at which point a node can no longer form ties. For the null model

simulation, the probability of dyad tie formation is proportional to the product of their degree deficits only.

To account for homophily in a node variable, the null simulation is modified so that the probability of tie formation also depends on the distance between the node variables. The variable-dependent probability is taken to fall off as a Gaussian function of the distance where a characteristic length scale l defines the preferred zone within which interactions are likely. The heterophily simulation is similar except that now interactions are more likely outside the zone defined by the characteristic length scale, which we refer to as the suppression length l_s . The purpose of these simulations is to see if a simple model of interactions including the variable of interest can minimize the error with the observed network at a well-defined value of the length scale. If so, additional evidence is thus provided for the operation of homophily or heterophily as well as an estimate of the length scale itself. For instance, if a heterophily simulation of ideology yielded a suppression length of $l_s = 2$ then that would imply that a group is more likely to fight with groups that are outside a distance equal to half the range of the full five-point ideological scale we deploy below. The appendix shows the simulations' mathematical formulation.

We augment these methods with two types of network regression analysis. We use the AMEN package in R to estimate an additive and multiplicative random effects model of militant infighting in Syria (Hoff, 2015). This model allows for the inclusion of both nodal and dyadic covariates. Latent space methods have been applied to international conflict data (Minhas, Hoff & Ward, 2016). We also use an ERGM method as a robustness check in the Online supplemental material.

The Syrian civil war

The Syrian civil war began as a peaceful Arab Spring movement but quickly formed into an armed insurgency against the regime of Bashar al-Assad (Lister, 2015). The conflict further evolved into a sectarian civil war and regional proxy conflict between the Gulf states and Iran (International Crisis Group, 2013). Non-state actors, including foreign fighters and transnational organizations like Al-Qaeda and Hezbollah, followed suit. Major powers – Russia and the United States – also intervened to shape conflict outcomes (Phillips, 2016).

Infighting data and variables

We collected data on 508 distinct infighting episodes between rebel groups in Syria from 1 January 2013 to

31 December 2015, yielding 697 *Infighting* dyadic ties (some episodes involved multiple groups and fronts, each of which was coded as a distinct infighting dyad). Our unit of analysis is the rebel group, which we define as a collection of armed fighters, ranging from several hundred to several thousand men and women, that have a commander and a distinct organizational identity as represented by a logo, and that uses violence in the course of a civil war or an insurgency to achieve publicly stated political aims against an incumbent regime and its allies. When rebel groups fight as part of formal coalitions or joint operation rooms, we disaggregate those broader units into their member groups and distribute infighting ties (dyads) to all the subgroups or to the specific ones involved in the infighting.

Rebel *Infighting*, our dependent variable, is defined as actual violent interactions between rebel groups. Violent interactions include armed clashes; firing artillery at rival positions; assassinating or executing rivals; arresting rebels or holding them captive; militarily advancing on a rival's territory or checkpoint with the intent of capturing it; and blowing up buildings, headquarters, or checkpoints that belong to one's rivals with car bombs or suicide attackers. *Infighting* does not include political disputes, defections, expulsions from the group, splintering, or counter-alliances. There were many infighting episodes that went on for days and weeks. Some turned into infighting campaigns that spanned several months. To accommodate such spans of continuous clashes, ties are defined at the month level with a tie between groups assigned for a given month if at least one violent interaction took place (for complete coding rules, see Online supplemental material). In the AME models described below, we use a square root transformation of infighting counts to approximate a normal distribution, with additional model specifications displayed in the Online supplemental material.

We selected 44 rebel groups to track based on think-tank and US government reports regarding the major players in the conflict. Out of these 44 groups, we analyze the 30 that were involved in at least one episode of infighting during the 2013–15 period. We define the *Power* variable to be the medium estimate of a group's number of members. These 30 groups range in *Power* from 500 to as many as 40,000 members (see Table I in the Online supplemental material, which also includes information on ideology, state sponsors, location, and years of existence). While the analysis set of 30 groups

Table I. Assortativity, community structure, and simulation results

Variable	Assortativity				Eigenvector correlation			Variable simulation	
	α	α_{null}	σ_{null}	p	EV	r	p	l_s/l	CI
2013–15									
N = 30, m = 697									
Conflict frame	–0.582(–)***	–0.265	0.036	<.0001	2MN	.454*	.012	2.5	(1.8,3.5)
Ideal polity	–0.342(–)***	–0.148	0.035	<.0001	2MN	.436*	.02	–	–
Territorial aspiration	–0.572(–)***	–0.287	0.035	<.0001	1MN	.394*	.03	2.2	(0.7,3.3)
Average ideology	–0.616(–)***	–0.284	0.036	<.0001	2MN	.409*	.02	1.7	(1.2,2.3)
Power	–0.308(–)**	–0.209	0.035	.005	2MN	.238	.21	2600	(1,000,4,000)
2014									
N = 22, m = 260									
Conflict frame	–0.529(–)***	–0.235	0.059	<.0001	2MN	.532*	.011	–	–
Ideal polity	–0.269(–)**	–0.126	0.055	.006	2MN	.455*	.03	–	–
Territorial aspiration	–0.540(–)***	–0.302	0.056	<.0001	2MN	.470*	.03	3.6	(2.1,5.2)
Average ideology	–0.554(–)***	–0.264	0.058	<.0001	2MN	.545**	.009	3.3	(1.6,5.3)
Power	–0.263(+)	–0.298	0.056	.54	2MP	.321	.14	–	–
2015									
N = 24, m = 424									
Conflict frame	–0.634(–)***	–0.290	0.045	<.0001	2MN	.442*	.03	1.8	(1.2,2.4)
Ideal polity	–0.390(–)***	–0.167	0.044	<.0001	2MN	.434*	.03	–	–
Territorial aspiration	–0.610(–)***	–0.291	0.045	<.0001	1MN	.465*	.02	1.4	(0.8,2.2)
Average ideology	–0.671(–)***	–0.307	0.044	<.0001	1MN	.431*	.04	1.4	(0.9,1.9)
Power	–0.326(–)**	–0.185	0.046	.003	2MN	.360	.08	–	–

Variables are displayed under the corresponding time periods (N = number of groups, m = number of ties). For assortativity: α is the assortativity of the observed network where the negative sign in parentheses indicates that α is less than α_{null} corresponding to heterophily (positive sign connotes homophily); α_{null} and σ_{null} are respectively the mean and standard deviation of the assortativity in the null simulation taken over 10,000 runs; the p-value p is the (two-tailed) fraction of runs exceeding $|\alpha - \alpha_{null}|$. For eigenvector correlation: EV is which one of the two most dominant eigenvectors has maximum correlation r with the variable (for heterophily, 1MN: most negative, 2MN: 2nd most negative; for homophily, 2MP: 2nd most positive); p is the p-value of r . For variable simulation: l_s is the mean suppression length (heterophily) and l is the mean interaction length (homophily) at which the minimum error occurs and CI is the 95% confidence interval (blank entries signify the absence of a clear minimum); 1,000 runs at each point (ranging from 0.1 to 6 in 0.1 increments for ideology variables; from 500 to 25,000 in increments of 500 for power) were used to generate 1,000 resamples of size 50 with replacement and then the l or l_s which minimized the squared error between the observed and simulated networks for each resample was found. * $p < .05$, ** $p < .01$, *** $p < .001$.

is far from exhaustive, we are confident that we have covered the major players involved in infighting.⁷

For Group *Ideology*, we hand-coded major ideological statements of the 30 groups that were involved in infighting episodes. We evaluated groups for three ideological areas of relevance to the Syrian conflict. Sectarianism serves as our *Conflict frame* variable: groups with high sectarianism scores cast the conflict as Sunnis vs. Shiites (Alawites), whereas groups with low sectarianism

scores have little or no anti-Shiite rhetoric. Salafism, which measures the extent to which groups ascribe to that highly puritanical strain of Sunni Islam, provides our *Ideal polity* variable. The use of Salafism better resolves differences within various stripes of Islamists than a simple secularism vs. Islamism scale. Revisionism is used for the *Territorial aspiration* component of ideology: groups with low scores seek to preserve Syria's territorial integrity, whereas a high score signifies a desire to abrogate it, in particular as do Caliphate-minded sectarian jihadists or Kurdish separatists.

A five-point scale was used for each component, and nationalists/Kurdish Separatists tend to fall on the low end (1) and sectarian jihadists on the high end (5) (see Table IV in the Online supplemental material for the ideological scores of the 30 groups). In addition, we also constructed an *Average ideology* variable from the average of *Conflict frame*, *Ideal polity*, and *Territorial aspiration* to

⁷ We compared our groups to the Uppsala Conflict Data Program (UCDP) dataset, which collected infighting episodes at the year level. We believe our dataset represents an important step forward because it is at the month level and is focused on individual groups rather than alliances of groups. Although here we limit our network to the 30 groups that engaged in infighting, we also substantiated our findings using all 44 groups through both AME and ERGM analyses (see Online supplemental material).

serve as an aggregate variable for visualization purposes and to provide an additional test of overall homophily.

Secular nationalists are represented by the Free Syrian Army (FSA), an umbrella organization that has many affiliated brigades. The FSA frames the Syrian rebellion as a national and democratic revolution that encompasses Syria's diverse ethnic and religious communities. It avoids overt sectarianism and rejects the goal of establishing an Islamic state ruled by strict religious laws (International Crisis Group, 2012).

On the other end of the ideological spectrum are groups like Al-Nusrah Front (ANF) and the Islamic State (ISIL). Formed in January 2012, ANF is an outwardly sectarian and jihadist faction that frames the conflict not as a revolution but rather as a religious war against a secular regime ruled by heretical Alawites. It calls for the establishment of an Islamic state governed by strict religious law. In 2013, ISIL broke up the ranks of the ANF to form an even more extreme sectarian jihadist faction. Its goal has been to carve out an Islamic state exclusively for Sunni Muslims that stretches from western Iraq to northeastern Syria.

Residing between the two poles of secular nationalist and sectarian jihadist are many Islamist factions ranging from Muslim Brotherhood affiliates such as the Al-Tawhid Brigade (ATB) to Salafists such as Ahrar al-Sham Islamic Movement (ASIM). We categorize these groups as Salafist nationalists because they want to establish an Islamic state within the extant boundaries of Syria's national territory and do not frame the conflict in overtly sectarian terms.

Kurdish communities formed their own combatant organizations, notably the People's Protection Units (Yekîneyên Parastina Gel, YPG), to safeguard their territories from both regime forces and hostile rebels (International Crisis Group, 2014). The secular YPG views Kurdish co-ethnics as its primary constituency, for which it seeks autonomy within, or separation from, the Syrian state.

Figure 1 displays the number of infighting incidents by ideological dyads across time. Infighting increased dramatically across the network in 2014 and 2015. Groups with an ideological difference have more infighting ties in every year and overall than groups with shared ideologies. Jihadist groups also fight among themselves more frequently than do dyads composed of secular nationalists or Kurdish separatists.

Rebel infighting appeared in every Syrian governorate, but the vast majority of the infighting took place in the rebel-held areas of Aleppo (38%) and the Damascus countryside (19%), followed by Idlib and Dayr al-Zawr

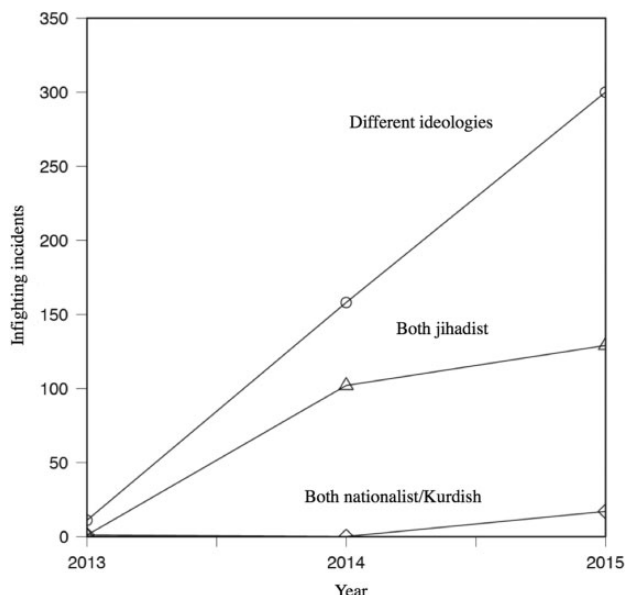


Figure 1. Infighting incidents by ideological dyads

Ideology is binarized: 3 or greater equals jihadist; lower than 3 equals nationalist/Kurdish factions.

(11% each). It is worth noting that most of Syria's oil and gas resources are concentrated in the eastern part of the country near the Iraqi border – that is, in Dayr al-Zawr and Hasaka. Rebel infighting in those regions is about 11% and 7%, respectively. Thus, most of the infighting took place outside of the resource-rich regions in the period under consideration.

ISIL fought the most with rival factions; it was involved in 41% of all infighting episodes, followed by ANF (15%), ASIM (8%), and Jaish al-Islam (JAI) and the Kurdish YPG (7% each). FSA-affiliated factions were involved in about 6% of the infighting episodes. Our dyadic analysis is undirected, meaning we do not distinguish between who began the hostilities and who was merely defending.

State sponsorship at the node-level (rebel group) is simply coded as 1 if the group had a state sponsor at any point during the period 2013–15. *State sponsorship* at the dyad level is coded as 1 if the two members of the dyad had any overlapping state sponsors. The major state sponsors of rebels have been Turkey, Saudi Arabia, Qatar, Jordan, and the United States (Phillips, 2016). Some groups have multiple sponsors. There are many organizations with no or unknown state sponsors. These include ISIL and ANF, but also many of the smaller groups.

We included two additional control variables for use in our network regression analysis. *Operational location*

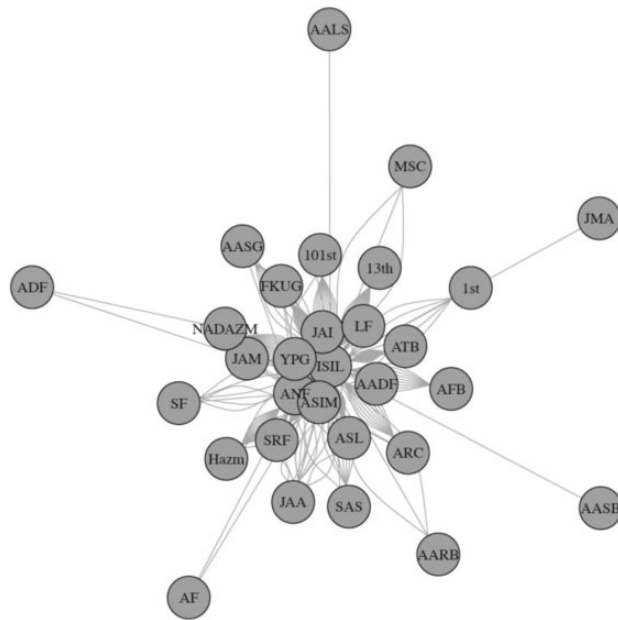


Figure 2. The network of Syrian militant infighting ties
A tie indicates a single infighting relationship with another group.

may be an important practical factor in driving infighting. Rebels that are in close proximity to each other can more easily fight than those who are far apart. *Islamic State* is a binary node level variable, coded 1 for ISIL and 0 if not. ISIL had the largest number of infighting ties in our network, and it is a highly ideological rebel group. We include this control to ensure ISIL was not the sole driver behind our ideology findings.

Network description

Figure 2 displays a visual representation of the network of infighting ties. Each line or tie denotes a single infighting episode. Observe that ISIL, along with ANF, ASIM, JAI, and YPG, form the center of the network, which means that they are the most frequent participants in interrebel conflicts (Table II in the Online supplemental material displays the descriptive statistics of that network and our variables).

Results

Assortativity and network simulation

A network whose elements correspond to the number of clashes between groups forms the basis of our analysis. Table I shows the observed assortativity values of the infighting network with respect to the ideology and power variables for the complete data time period (2013–15). We discuss the ideology results first.

Considering 2013–15, the observed assortativity for every ideology variable is more negative than its assortativity in the null simulation, indicating heterophily. All three ideology components and their averages are found to have highly significant deviations from the null assortativity. They also have significant correlations with one or the other of the two eigenvectors with the most negative eigenvalues. This alignment with salient structural features in the network suggests that ideological heterophily plays an important role in shaping patterns of conflict. The 2014 and 2015 assortativity results show a similar pattern to 2013–15 with all the ideology variables again significant. For 2014, all of the components best correlate with the second most negative eigenvector. However, for 2015, *Territorial aspiration* best correlates with the most negative eigenvector, a change which suggests that, as concerns over the disintegration of Syria grew, the salience of *Territorial aspiration* intensified. Overall, we conclude that the high significance of the ideology assortativity for the full data and both individual years provides strong support for Hypothesis 2, that the likelihood of infighting increases with the ideological distance between groups.

As all ideology variables display heterophily, the suppression lengths, within which infighting is less likely, are reported for the ideology simulations with variable-based interactions. For the 2013–15 data period, well-defined suppression lengths for the heterophily simulation are found for *Territorial aspiration*, *Conflict frame*, and *Average ideology*. That $l_5 = 1.7$ for *Average ideology* indicates that the probability of infighting becomes substantially larger when the ideological distance between the groups in a dyad exceeds about half the full ideology scale. Note that *Ideal polity* does not have a well-defined suppression length consistent with its relatively small magnitude assortativity. The 2015 time period is similar to the full dataset, but the suppression lengths are shorter indicating a narrowing of the ideological zone for which infighting is substantially less probable. For 2014, however, only *Territorial aspiration* and *Average ideology* exhibit well-defined suppression lengths although at relatively large values over 3. The difference between 2014 and 2015 parallels the decrease in assortativity to more negative values in the latter period.

Figure 3 visualizes the infighting network using the two least-ranked eigenvectors as node coordinates. The nodes are shaded with respect to their *Average ideology* scores. The dominant pattern is represented by the most negative eigenvector and shown on the vertical axis. It essentially corresponds to ISIL arrayed against everyone else. It is the second most negative eigenvector, shown

Table II. AME regression results

	Model 1	Model 2	Model 3	Model 4	Model 5
Beta values	Power	Ideology	Sponsor	Full model, all groups	Full model, cont. for ISIL
Intercept	-0.054 (0.084)	-0.063 (0.082)	-0.01 (0.082)	-0.082 (0.298)	-0.200 (0.328)
<i>Node-level variables</i>					
Average ideology (1 to 5)				-0.003 (0.042)	-0.002 (0.005)
Power (0.5–40)				-0.005 (0.005)	-0.002 (0.005)
Sponsorship (Y/N)				-0.003 (0.116)	0.030 (0.125)
ISIL (Y/N)					1.449** (0.557)
<i>Dyad-level variables</i>					
Ideological difference (0–4)		0.035*** (0.010)		0.034*** (0.010)	0.033** (0.011)
Power difference (0–39.5)	0.004* (0.002)			0.005* (0.002)	0.005* (0.002)
Sponsor overlap (Y/N)			-0.03 (0.025)	-0.005 (0.025)	-0.009 (0.026)
Location overlap (Y/N)				0.067* (0.032)	0.071* (0.034)
<i>Variance parameters</i>					
Pmean (va)	0.044	0.042	0.042	0.067	0.050
Pmean (ve)	0.027	0.027	0.027	0.246	0.028
psd (va)	0.013	0.012	0.012	0.021	0.016
psd (ve)	0.001	0.001	0.001	0.012	0.002
N	30	30	30	30	30

The power range is in units of a thousand. DV is the square root of the count of infighting incidents. * $p < .05$, ** $p < .01$, *** $p < .001$.

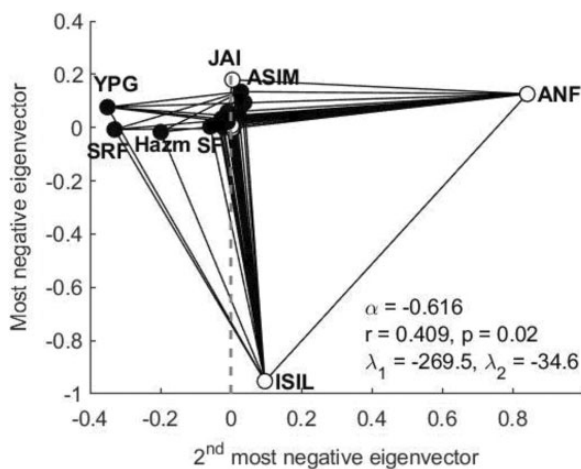


Figure 3. The infighting network structure for the 2013–15 period

Solid circles indicate groups with *Average ideology* < 3 (open circles ≥ 3). Links between groups indicate at least one clash. Assortativity (α) value of *Average ideology* and correlation (r) and p -value with 2nd most negative eigenvector shown. 1st and 2nd most negative eigenvalues denoted by λ_1 and λ_2 . Vertical dashed line marks division into communities.

on the horizontal axis, that best correlates with *Average ideology*. This eigenvector indicates that *Average ideology* tends to pit sectarian jihadists (ANF, ISIL) on the right

against secular groups (YPG, Hazm) on the left. Powerful Salafist groups (ASIM, JAI) are found in the middle. Although the groups on the left show diversity in *Territorial aspiration*, they are more uniform with respect to *Conflict frame* and *Ideal polity*, being less sectarian and less Islamist than the jihadists on the right – an observation also consistent with those ideology variables correlating with the second most negative eigenvector, while *Territorial aspiration* best correlates with the most negative eigenvector. The left grouping is anchored by secular, non-sectarian groups (Syrian Revolutionaries Front and YPG) at the extreme whereas the highly sectarian ANF anchors the jihadist side.

Turning to *Power*, the difference between the observed *Power* assortativity and the null mean value for 2013–15 is negative, indicating heterophily, and also highly significant, thereby supporting the asymmetrical power Hypothesis 1a that strength disparity tends to increase infighting. Although *Power* does not correlate well with either of the two most negative eigenvectors, the variable-based simulation does yield a well-defined suppression length of 2,600. Heterophily with respect to power is also indicated by the significant negative assortativity deviation in 2015. The 2014 period, however, shows no significant effect of power and it is greater than the null assortativity indicating a tendency toward

homophily rather than heterophily. Consequently, we conclude that there is strong evidence for the asymmetrical power dynamic although it is less consistent across analysis types and time than ideology.

Network regression results

Table II displays the results of network regression analysis. Hypotheses 1a and 1b are evaluated through examining a difference in *Power*. A positive and statistically significant relationship with *Infighting* would provide evidence in favor of Hypothesis 1a, power asymmetry: as the difference in size between two groups grows, they become more likely to fight. Conversely, a negative and statistically significant relationship would demonstrate support for Hypothesis 1b, power symmetry.

We see evidence for Hypothesis 1a in Table II: difference in size between groups generally increases the likelihood of *Infighting*. We note that while this finding has the largest substantive impact, it does not achieve statistical significance in all models run: in binary outcome variables and a raw count, it does not achieve statistical significance at all (see Online supplemental material).

Hypothesis 2 is evaluated using the *Ideological distance* variable, defined as the difference between *Average ideology* values of the groups in a dyad. Table II shows that *Ideological distance* displays a positive and statistically significant relationship regardless of which other variables at the dyad or node level are included, thereby indicating a greater tendency for infighting among ideologically dissimilar groups. This variable is by far the most consistent in terms of achieving statistical significance (regardless of model used, see Online supplemental material) and has a relatively large substantive impact (as compared with, say, location or sponsorship), lending strong support to Hypothesis 2. It even holds in terms of direction and significance of beta values in Model 5, which controls for ISIL at the node level. ISIL, the most frequent participator in infighting, has a strong effect on the network: this is manifest by the most negative eigenvector as shown in Figure 3, and by a large effect size and statistically significant result in Model 5. Thus, even when controlling for ISIL, ideology still has a statistically significant effect on the likelihood of infighting.

Hypothesis 3 is evaluated using shared *State sponsorship*. A negative and statistically significant relationship between *State sponsorship* and *Infighting* would provide evidence in favor of Hypothesis 3 – that *Shared sponsorship* makes infighting less likely. We find no evidence of a relationship between *Shared sponsorship* at the dyad level

and *Infighting* (even when controlling for having or lacking a sponsor at the node level). The absence of a relationship could reflect the limited control that state sponsors have over their clients or a deliberate strategy to hedge by betting on multiple groups without concern over their collective cohesiveness. Rubin (2002: 198) noted this pattern in the case of Pakistan's support for the Afghan *mujahidin* during their anti-Soviet jihad; and Staniland (2014: 163) in the case of India's support of Tamil factions in Sri Lanka during the 1980s.

ERGM results in the Online supplemental material confirm our findings concerning group size and ideology (though as with the AME analysis, power is less consistent across model specifications). Together, these three methods provide robust support for the ideological distance hypothesis, and support for power asymmetry shaping Syria's interrebel conflicts. We also find evidence that shared location makes infighting more likely.

Discussion and conclusions

We addressed the puzzle of interrebel wars at the dyadic level and tested it using data from the Syrian civil war. The results of our analyses using assortativity, community structure detection, network simulation, and AME regression indicate that two mechanisms – *Ideological distance* and *Power asymmetry* – predict which rebel dyads are at most risk for infighting. Ideologically opposed groups have a higher propensity for infighting than ideologically proximate ones, and groups of disparate strength are more likely to fight with each other than groups of comparable power. These findings suggest that power and ideology need not be viewed as competing explanations of rebel infighting. Instead, they complement one another. For example, power may explain *why* rebels engage in fratricide, including the quest for security and hegemony in a competitive landscape, and our analysis demonstrates that they are more likely to engage in war with groups of dissimilar size. Ideology, though, appears to tell us *with whom* groups are likely to fight in order to achieve their power aims. The greater the ideological distance between groups, the more likely they are to fight one another.

Our ideology findings may raise an endogeneity challenge. It could be asserted that infighting driven by power considerations compels groups to accentuate their ideological divides to justify or motivate the conflict. In this scenario, conflictual relationships drive the ideological distance exhibited in infighting networks, not the other way around. This objection assumes that militant groups arise as ideological blank slates, contrary to the

fact that the founders of such groups often have strong ideological orientations from the outset. One of the key insights from Staniland's (2014: 33) social institutional theory of rebellion is that insurgents 'draw upon prewar political life in order to quickly form organizations that can handle the strains of violence'. Many of the individuals who would go on to form Syria's major Islamist rebel groups were actually in jail at the start of the revolution due to their prior Islamist activism and then subsequently released (Lister, 2015: 53–55).

Additionally, ideological manifestos and political programs, an important element of our coding, are typically issued by groups shortly after their formation. Their ideological statements, therefore, are biased toward a time before these groups have entered into interrebel wars. In other words, a variable biased toward the earlier period in the data cannot significantly correspond to patterns of conflict over two distinct years, 2014 and 2015, as well as the entire time period and yet be causally dependent on the infighting ties which are skewed toward a later period in the data (see Table I). Table I also illustrates that ideology corresponded to the second most negative eigenvector, indicating that ideological differences are a greater force than power asymmetries in shaping the overall pattern of the network. Furthermore, ideology cannot be epiphenomenal to power since they are simultaneously significant in Models 4 and 5 of the AME regression analysis – that is, when we control for power, ideology is still statistically significant and remains significant in all models. Nor can ideology simply be a function of overlapping state sponsorship since no significant relationship was found between the latter and the infighting network. Finally, ISIL is not driving the relationship as this finding exists even when controlling for this group in Model 5.

Another possible challenge to our results stems from our use of militant claims of infighting. It can be argued that militants preferentially reveal clashes with ideologically distant groups and conceal the ones with their ideological kin, and thus our data underrepresents infighting among ideological brothers. Of course, such self-censorship may occur, but to raise this objection beyond conjecture, one must estimate its frequency, a difficult task given the lack of ground truth. The use of militant claims enables construction of a dataset that is large enough to employ network analysis for a single conflict, an enterprise that can be pursued much less robustly using standard conflict datasets like UCDP. Although comparison with the UCDP data is difficult, when searching for missing infighting dyads (here, meaning pairs that fight at least once) among our ten

most prominent groups (by degree), the UCDP data provide no evidence of a missing dyad. If the concealment of infighting between ideologically proximate groups was frequent, then one would expect that at least one such pair would be found in UCDP that is absent from our data. That there is none leaves little evidentiary basis to support the self-censorship objection. Furthermore, our data reflect many fighting episodes between ISIL and ANF, two ideologically similar movements. This suggests that the presumption that rebels mask their infighting when it is politically inconvenient is contradicted repeatedly in at least one prominent conflict dyad.

Lastly, one might object that our analysis leaves out the state, an important factor in shaping interrebel conflicts. The state's accommodative arrangements can give some rebel groups a greater degree of freedom to attack their rivals because they are less concerned about fighting the state. For example, the Syrian regime has been accused of deliberately neglecting ISIL in its targeting policy so that a tacit alliance between the state and ISIL allowed the latter to concentrate its fighting resources on wiping out its rivals. Although accounting for selective targeting of rebel groups by the state would add to the substantive analysis of infighting, doing so is empirically difficult as the Syrian regime typically made claims of attacking 'terrorists' rather than specific groups.

We assess the impact of omitting the state through consideration of ISIL's effect on the network. Given that ISIL is responsible for 41% of all the infighting ties, the potential collusion between the Assad regime and ISIL would be the most consequential confounding factor arising from the omission of the state. However, Figure 3 shows that the second most negative eigenvector best correlates with *Average ideology* in the full data period and not the most negative eigenvector, corresponding to ISIL arrayed against all the other groups – a result suggesting that ISIL is not the dominant driver of our ideological distance finding. Furthermore, removing ISIL from the network yields the best correlation with the most negative eigenvector ($p = .02$), and the assortativity test still strongly supports ideological heterophily ($p < .0001$). That ideological heterophily persists in the pattern of infighting after removing ISIL provides strong evidence that infighting cannot be attributed to selective targeting by the state or its collusion with particular groups.

Our empirical findings in the Syrian case add weight to a burgeoning body of scholarship that makes the case that 'ideological considerations play a prominent role in guiding insurgent decisionmaking' (Hirose, Imai & Lyall, 2017: 48). This literature has empirically demonstrated how ideological variables can explain important

conflict processes in civil wars, such as anti-civilian atrocities (Straus, 2015) and mobilization effectiveness (Ugarriza & Craig, 2012; Costalli & Ruggeri, 2015). Our article shows that ideology also matters to rebel infighting and gives added credence to those who call for bringing armed politics back into the study of civil conflicts (Staniland, 2015; Balcells, 2017). Allowing for both power and ideology to impact insurgent factional dynamics should improve our understanding and anticipation of conflict trajectories in fragmented civil wars.

Replication data

The dataset, computational scripts, and Online supplemental material can be found at <http://www.prio.org/jpr/datasets>.

Acknowledgments

We thank Naazneen Barma, Cassy Dorff, Peter Krause, and Zane Kelly.

Funding

This research was funded by the Office of Naval Research under grants N00014-15-1-2549 and N00014-16-1-2919.

ORCID iD

Michael Gabbay  <https://orcid.org/0000-0002-4116-6782>

References

- Akcinaroglu, Seden (2012) Rebel interdependencies and civil war outcomes. *Journal of Conflict Resolution* 56(5): 879–903.
- Bakke, Kristin; Kathleen Gallagher Cunningham & Lee JM Seymour (2012) A plague of initials: Fragmentation, cohesion, and infighting in civil wars. *Perspectives on Politics* 10(2): 265–283.
- Balcells, Laia (2017) *Rivalry and Revenge: The Politics of Violence during Civil War*. New York: Cambridge University Press.
- Bapat, Navin & Kanisha D Bond (2012) Alliances between militant groups. *British Journal of Political Science* 42(4): 793–824.
- Christia, Fotini (2012) *Alliance Formation in Civil Wars*. New York: Cambridge University Press.
- Costalli, Stefano & Andrea Ruggeri (2015) Indignation, ideologies, and armed mobilization: Civil war in Italy, 1943–45. *International Security* 40(2): 119–157.
- Cunningham, Kathleen Gallagher; Kristin M Bakke & Lee JM Seymour (2012) Shirts today, skins tomorrow: Dual contests and the effects of fragmentation in self-determination disputes. *Journal of Conflict Resolution* 56(1): 67–93.
- Desmarais, Bruce A & Skyler J Cranmer (2017) Statistical inference in political networks research. In: Jennifer Nicoll Victor, Alexander H Montgomery & Mark Lubell (eds) *The Oxford Handbook of Political Networks*. New York: Oxford University Press, 203–220.
- Dorussen, Han; Erik A Gartzke & Oliver Westerwinter (2016) Networked international politics: Complex interdependence and the diffusion of conflict and peace. *Journal of Peace Research* 53(3): 283–291.
- Fjelde, Hanne & Desirée Nilsson (2012) Rebels against rebels: Explaining violence between rebel groups. *Journal of Conflict Resolution* 56(4): 604–628.
- Griffiths, Ryan D (2016) *Age of Secession: The International and Domestic Determinants of State Birth*. New York: Cambridge University Press.
- Gutiérrez Sanín, Francisco & Elisabeth Jean Wood (2014) Ideology in civil war: Instrumental adoption and beyond. *Journal of Peace Research* 51(2): 213–226.
- Hafner-Burton, Emilie M; Miles Kahler & Alexander H Montgomery (2009) Network analysis for international relations. *International Organization* 63(3): 559–592.
- Hirose, Kentaro; Kosuke Imai & Jason Lyall (2017) Can civilian attitudes predict insurgent violence? Ideology and insurgent tactical choice in civil war. *Journal of Peace Research* 54(1): 47–63.
- Hoff, Peter D (2015) Dyadic data analysis with AMEN (<https://arxiv.org/abs/1506.08237>).
- Hoover Green, Amelia (2016) The commander's dilemma: Creating and controlling armed group violence. *Journal of Peace Research* 53(5): 619–632.
- International Crisis Group (2012) Tentative Jihad: Syria's fundamentalist opposition (<https://d2071andvip0wj.cloudfront.net/131-tentative-jihad-syria-s-fundamentalist-opposition.pdf>).
- International Crisis Group (2013) Syria's metastasising conflicts (<https://www.crisisgroup.org/middle-east-north-africa/eastern-mediterranean/syria/syria-s-metastasising-conflicts>).
- International Crisis Group (2014) Flight of Icarus? The PYD's precarious rise in Syria (<https://www.crisisgroup.org/middle-east-north-africa/eastern-mediterranean/syria/flight-icarus-pyd-s-precarious-rise-syria>).
- Jones, Seth (2017) *Waging Insurgent Warfare: Lessons from the Vietcong to the Islamic State*. New York: Oxford University Press.
- Krause, Peter (2013/14) The structure of success: How the internal distribution of power drives armed group behavior and national movement effectiveness. *International Security* 38(3): 72–116.
- Krause, Peter (2017) *Rebel Power: Why National Movements Compete, Fight, and Win*. Ithaca, NY: Cornell University Press.

- Kydd, Andrew & Barbara F Walter (2002) Sabotaging the peace: The politics of extremist violence. *International Organization* 56(2): 263–296.
- Lichbach, Mark Irving (1995) *The Rebel's Dilemma*. Ann Arbor, MI: University of Michigan Press.
- Lister, Charles R (2015) *The Syrian Jihad: Al-Qaeda, the Islamic State, and the Evolution of an Insurgency*. New York: Oxford University Press.
- Maoz, Zeev (2010) *Networks of Nations: The Evolution, Structure, and Impact of International Networks, 1816–2001*. New York: Cambridge University Press.
- McLauchlin, Theodore & Wendy Pearlman (2012) Out-group conflict, in-group unity? Exploring the effect of repression on intramovement cooperation. *Journal of Conflict Resolution* 56(1): 41–66.
- Minhas, Shahryar; Peter D Hoff & Michael D Ward (2016) A new approach to analyzing coevolving longitudinal networks in international relations. *Journal of Peace Research* 53(3): 491–505.
- Newman, Mark EJ (2006) Finding community structure in networks using the eigenvectors of matrices. *Physical Review E* 74(3): 1–22.
- Oppenheim, Benjamin; Abbey Steele, Juan Vargas & Michael Weintraub (2015) True believers, deserters, and traitors: Who leaves insurgent groups and why. *Journal of Conflict Resolution* 59(5): 794–823.
- Phillips, Christopher (2016) *The Battle for Syria: International Rivalry in the New Middle East*. New Haven, CT: Yale University Press.
- Pischedda, Costantino (2018) Wars within wars: How windows of opportunity and vulnerability cause inter-rebel fighting. *International Security* 43(1): 138–176.
- Rubin, Barnett R (2002) *The Fragmentation of Afghanistan*. New Haven, CT: Yale University Press.
- Salehyan, Idean (2010) The delegation of war to rebel organizations. *Journal of Conflict Resolution* 54(3): 493–515.
- Salehyan, Idean; David Siroky & Reed M Wood (2014) External rebel sponsorship and civilian abuse: A principal–agent analysis of wartime atrocities. *International Organization* 68(3): 633–661.
- Seymour, Lee (2014) Why factions switch sides in civil wars: Rivalry, patronage, and realignment in Sudan. *International Security* 39(2): 92–131.
- Seymour, Lee; Kristin M Bakke & Kathleen Gallagher Cunningham (2016) E pluribus unum, ex uno plures: Competition, violence, and fragmentation in ethnopolitical movements. *Journal of Peace Research* 53(1): 3–18.
- Staniland, Paul (2014) *Networks of Rebellion: Explaining Insurgent Cohesion and Collapse*. Ithaca, NY: Cornell University Press.
- Staniland, Paul (2015) Militias, ideology, and the state. *Journal of Conflict Resolution* 59(5): 770–793.
- Straus, Scott (2015) *Making and Unmaking Nations: War, Leadership and Genocide in Modern Africa*. Ithaca, NY: Cornell University Press.
- Tamm, Henning (2016) Rebel leaders, internal rivals, and external resources: How state sponsors affect insurgent cohesion. *International Studies Quarterly* 60(4): 599–610.
- Ugarriza, Juan & Matthew J Craig (2012) The relevance of ideology to contemporary armed conflicts: A quantitative analysis of former combatants in Colombia. *Journal of Conflict Resolution* 57(3): 445–477.
- Ward, Michael; Katherine Stovel & Audrey Sacks (2011) Network analysis and political science. *Annual Review of Political Science* 14: 245–264.
- Zech, Steven & Michael Gabbay (2016) Social network analysis in the study of terrorism and insurgency: From organization to politics. *International Studies Review* 18(2): 214–243.

EMILY KALAH GADE, b. 1985, PhD in Political Science (University of Washington, 2017); Acting Assistant Professor of Political Science and Moore/Sloan Innovations in Data Science Postdoctoral Fellow, University of Washington (2017–); research area: political violence, data science.

MOHAMMED M HAFEZ, b. 1970, PhD in International Relations (London School of Economics, 2000); Associate Professor of National Security Affairs, Naval Postgraduate School (2008–); research area: Islamist political violence.

MICHAEL GABBAY, b. 1963, PhD in Physics (University of Chicago, 1997); Senior Principal Physicist, Applied Physics Laboratory, University of Washington (2009–); research area: political networks.

Appendix

We describe the formalism used to calculate the assortativity, relate it to the eigenvector spectrum of the modularity matrix, and formalize the network simulation. We consider a symmetric network with N nodes represented by the adjacency matrix with components A_{ij} , equal to the number of ties between nodes i and j ($A_{ij} = A_{ji}$). The degree of node i is the sum of its ties, $k_i = \sum_{j=1}^N A_{ij}$. The total number of ties in the network is m . These network quantities are used to define the components of the modularity matrix $\mathbf{B}$,

$$B_{ij} = A_{ij} - \frac{k_i k_j}{2m}, \quad (1)$$

which is the difference between the observed tie strength and what would be expected from a null process in which ties are formed in proportion to the product of node degrees without regard to any interactions driven by node variables. Assortativity is the standard measure used to assess homophily of network tie formation with

respect to a scalar variable (as we operationalize power and ideology).

The assortativity that the network displays with respect to a node variable x can be related to the modularity matrix via

$$\alpha = \frac{\sum_{i,j} B_{ij} x_i x_j}{\sum_{i,j} A_{ij} x_i^2 - (k_i k_j / 2m) x_i x_j}, \quad (2)$$

which is equivalent to the correlation of x over ties (Newman, 2006).

The spectrum of the modularity matrix is defined via $\mathbf{B}\mathbf{u}_\nu = \lambda_\nu \mathbf{u}_\nu$ where λ_ν is the eigenvalue corresponding to eigenvector $\mathbf{u}_\nu$ with eigenvectors indexed in order of decreasing eigenvalue. Newman (2006) shows that the assortativity can be expanded as a sum of the modularity matrix eigenvalues where the weight associated with each λ_ν is proportional to the square of the inner product of the vector formed by the x_i and the associated eigenvector $\mathbf{u}_\nu$, $\left(\sum_{i=1}^N u_{\nu i} x_i\right)^2$. Homophily, therefore, can be manifested by significant correlations of the variable with a highly ranked eigenvector (of positive eigenvalue), not just the leading one. Heterophily, on the other hand, is

manifested by significant correlations with low-ranked eigenvectors, i.e., those with the most negative eigenvalues.

To formulate the null and variable-based simulations, we denote the maximum possible degree of node i by D_i (its degree in the empirical network) and its degree at iteration step n by $k_i(n)$. The deficit between the maximum and current degrees is then $D_i - k_i(n)$. Accordingly, the probability of tie formation between nodes i and j at step n is

$$p_{ij}(n) = \begin{cases} K_n (D_i - k_i(n))(D_j - k_j(n)) f(x_i - x_j), & i < j \\ 0, & i \geq j. \end{cases} \quad (3)$$

K_n is a normalization constant so that $\sum_{i,j} p_{ij}(n) = 1$. The bottom line prevents self-ties (the $i > j$ case need not be considered separately because the network is symmetric). The function $f(x_i - x_j)$ accounts for the dependence of the probability upon the difference of the variable x between the two nodes. We make three choices for $f(x_i - x_j)$ corresponding to: (1) the null simulation, $f(x_i - x_j) = 1$; (2) the homophily simulation, $f(x_i - x_j) = \exp(-0.5(x_i - x_j)^2 / l^2)$ (a Gaussian); and (3) the heterophily simulation, $f(x_i - x_j) = 2 - \exp(-0.5(x_i - x_j)^2 / l_s^2)$.

Appendix 2

Networks of Cooperation: Rebel Alliances in Fragmented Civil Wars

Networks of Cooperation: Rebel Alliances in Fragmented Civil Wars

Journal of Conflict Resolution
2019, Vol. 63(9) 2071-2097
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0022002719826234
journals.sagepub.com/home/jcr



Emily Kalah Gade¹, Michael Gabbay²,
Mohammed M. Hafez³, and Zane Kelly²

Abstract

When rebels make alliances, what informs their choice of allies? Civil wars are rarely simple contests between rebels and incumbent regimes. Rather, rival militant networks provide the context in which these fragmented conflicts unfold. Alliances that emerge within this competitive landscape have the power to alter conflict trajectories and shape their outcomes. Yet patterns of interrebel cooperation are understudied. The existing scholarship on rebel alliances focuses on why rebels cooperate, but little attention is given to the composition of those alliances: with whom rebels cooperate. We explore how power, ideology, and state sponsorship can shape alliance choices in multiparty civil wars. Employing network analysis and an original data set of tactical cooperation among Syrian rebels, we find compelling evidence that ideological homophily is a primary driver of rebel collaboration. Our findings contribute to an emerging literature that reasserts the role of ideology in conflict processes.

Keywords

Syria, civil war, fragmentation, alliances, social network analysis, ideology

¹Department of Political Science, University of Washington, Seattle, WA, USA

²Applied Physics Laboratory, University of Washington, Seattle, WA, USA

³Department of National Security Affairs, Naval Postgraduate School, Monterey, CA, USA

Corresponding Author:

Emily Kalah Gade, Department of Political Science, University of Washington, Gowen Hall, Seattle, WA 98195, USA.

Email: ekgade@uw.edu

Civil wars are rarely simple contests between unified rebels and incumbent regimes. Instead, they usually feature divided rebel movements with multiple factions competing over leadership, territory, resources, and fighters (Bakke, Cunningham, and Seymour 2012). Forging unity among armed groups is a challenge because credible commitment problems make binding obligations difficult to initiate and sustain. Cooperation also involves trade-offs between enhancing one's power capabilities and decision-making autonomy, which may incline some rebels to forgo alliances that diminish their independence. Ideological considerations also affect rebel cooperation: factions that harbor competing visions for the future are likely to view alliances with rivals as short-lived exigencies at best. Yet despite these barriers, cooperation among armed factions does occur. Between 1946 and 2008, 181 of the 345 groups in civil wars, more than 52 percent, "have initiated positive associations with each other while fighting with the government" (Akcinaroglu 2012, 890).

The prevalence of rebel cooperation alongside competition generates two puzzles: why do rebels cooperate and with whom do they cooperate? Literature on rebel cooperation has focused on the *why* question. Interrebel alliances emerge between factions seeking to augment their capabilities and improve their tactical productivity (Lichbach 1995), balance against their rivals through minimum winning coalitions (MWCs; Christia 2012), and increase their overall odds of victory by institutionalizing joint command and control of military operations (Akcinaroglu 2012). Little is known, however, about the factors that shape the composition of rebel alliances, that is, *with whom* rebels cooperate. Civil wars can involve hundreds of rebel brigades, which could produce countless cooperative alignments. This translates into rebels having choices when pursuing cooperation to achieve their conflict objectives. What explains their choice of allies?

In addressing this puzzle, we make distinct theoretical, methodological, and empirical contributions. Theoretically, we explore three logics of alliance composition related to ideology, power, and state sponsorship and make predictions about how they might shape militant collaboration. We posit that ideological proximity in rebel networks should yield greater militant cooperation than ideological distance, thus challenging the prevailing assumption that ideology is a minor consideration in alliance formation (Christia 2012). We operationalize ideology in civil wars along three dimensions—*conflict framing*, *conception of the ideal polity*, and *territorial aspiration*—and show that agreement within those issue areas facilitates cooperation among rebel factions. Through conflict framing, a rebel group identifies whom it is primarily fighting for and against, casting both in-group and out-groups with respect to its preferred cleavage whether ethnic, religious, economic, or political. A group's conception of the ideal polity identifies its vision for the postconflict social and political order and its territorial aspiration identifies the boundaries of this future order. Unpacking ideology into these distinct dimensions allows for a more nuanced understanding of factional alignments than the classification of rebels into broad categories such as nationalists, separatists, socialists, and fundamentalists.

We complement our ideological analysis with a thorough consideration of how power and state sponsorship inform alliance choices. For power, we propose contrasting hypotheses of symmetrical and asymmetrical alliance formation. We posit that an overriding concern for capability aggregation in rebel movements will tend to produce *symmetric alliances* (cooperation between groups of comparable strength), whereas the desire of strong groups to form alliances that maximize decision-making autonomy vis-à-vis rivals will generate *asymmetric alliances* (cooperation between groups of dissimilar power capabilities). As for state sponsorship in rebel alliances, we test the hypothesis that rebel groups that share the same state sponsor will cooperate more frequently than those with no overlapping external sponsors.

Methodologically, we employ social network analysis to yield insights into rebel cooperation within fragmented conflicts. Network approaches are widespread in political science, yet few have sought to apply them to multiparty civil wars (Zech and Gabbay 2016; Metternich et al. 2013). Research on civil conflicts calls for a network approach because rebel groups do not make choices to align with others in a vacuum, but rather their choices are likely to hinge on the alliance preferences of the other groups in the rebel movement. Thus, social network analysis can better capture the theoretical patterns we would expect to observe than the standard statistical assumption of independence of observations when examining dyads in multiparty civil wars. We use additive and multiplicative effects (AME) models to evaluate the relationships between our three proposed variables simultaneously. In addition, as a robustness check, we use simulations of network tie formation to augment these findings.

Empirically, we test our hypotheses as they relate to factional cooperation in Syria's civil war. One of the world's bloodiest conflicts, the Syrian rebellion features a complex set of actors with local, regional, and international ties. We use primary insurgent sources, including more than 9,000 unique claims of attacks, to construct an original data set for more than 220 insurgent groups active since the onset of conflict through mid-2015. We form a network of militant tactical cooperation from claims of joint operations and investigate its structure with respect to ideology as obtained from manual coding of primary source materials, power as measured by group size, and state sponsorship as reported in informed secondary sources. We find compelling evidence that ideological homophily is a driver of rebel cooperation in Syrian militant networks. We also find some evidence in favor of symmetric alliances rather than asymmetric ones, but it is inconsistent across our analysis. However, we do not find support for the proposition that overlapping state sponsorship in rebel dyads increases cooperation.

The question of alliance composition has important strategic implications. Understanding the dynamics of rebel cooperation can yield policy insights for prompting or dissuading alliances between rebel groups in multifactional civil wars. Recent conflicts illustrate vividly how the composition of rebel alliances can shape conflict trajectories in dramatic ways. In 2007, American-led coalition forces in Iraq successfully exploited rifts in rebel unity to turn the tide in the war. Nationalist

insurgents became increasingly alienated by Al-Qaeda in Iraq, their former jihadist ally, and as such were willing to switch sides to the American coalition. The dissolution of the nationalist–jihadist alliance contributed to a substantial reduction in violence, until the resurgence of the Islamic State six years later.

In contrast, the United States viewed with concern the fragmentation of Syria's rebel movement, preferring a unified and cohesive rebel alliance composed of moderate rebel factions that could topple the Assad regime. The United States, however, could not overlook the presence of extreme Islamist factions in the rebel movement and thus deprioritized the objective of rebel unity. Ultimately, it limited its support to a narrow sector of acceptable militant groups, which proved ineffective against an incumbent regime backed by a unified and powerful coalition consisting of Iran, Hezbollah, and Russia.

Beyond these recent examples, the literature on rebel fragmentation points out that (dis)unity has important implications for several other conflict processes. Civil wars with divided rebel factions last longer, are more violent, and have higher rates of recurrence than wars with unified rebel movements (Cunningham, Bakke, and Seymour 2012; Wood and Kathman 2015; Driscoll 2012; Cunningham 2013; Rudloff and Findley 2016). Conversely, movements led by a hegemonic faction are more likely to be successful than more diffuse movements (Krause 2017). By illuminating the drivers of rebel cooperation, this study, therefore, makes a contribution to understanding a dynamic of great consequence in fragmented conflicts.

Ideology in Rebel Alliances

Rebel groups have political preferences and moral visions for which they are fighting. The preceding century has highlighted the capacity of Marxist, nationalist, fundamentalist, and fascist ideologies to mobilize millions of people for revolutions, insurgencies, civil wars, and genocide. Although not all civil conflicts are driven by ideological divides and not all rebels are motivated by ideological considerations, diversity of political demands typify fragmented civil conflicts, which are the most common form of wars today (Seymour, Bakke, and Cunningham 2016, 5, 6; Jones 2017, 168).¹

Recent scholarship has rediscovered the critical role that ideology plays in conflict processes (Ugarriza and Craig 2012; Costalli and Ruggeri 2015; Staniland 2015; Balcells 2017). Ideology is a source of collective identity and can help forge group cohesion in the context of civil wars by orienting commanders and foot soldiers toward a clear set of objectives (Gutiérrez Sanín and Wood 2014). It can also motivate commitment and sacrifice, remove inhibitions to violence, and reprioritize collective incentives above self-regarding considerations (Lichbach 1995, 92, 93; Walter 2017, 19, 20; Kim 2018, 308). Additionally, ideological socialization has been shown to improve battlefield discipline and dissuade defections to the state (Oppenheim et al. 2015; Hoover Green 2016).

We contribute to this burgeoning literature by proposing mechanisms that link ideology to the choice of allies in rebel movements, thus challenging the prevailing

assumption in the literature that ideology is a secondary consideration in alliance formation. We expect these mechanisms to apply to the spectrum of cooperative relationships ranging from joint operations at the tactical level to formal alliances at the strategic level. Joint operations (our empirical measure) consist of two or more rebel groups conducting an attack together (Bapat and Bond 2012, 19). We focus here on tactical collaboration because only 17.6 percent of rebel cooperation between 1946 and 2008 was at the level of formal alliances (Akcinaroglu 2012, 890).

Joining Gutiérrez Sanín and Wood (2014, 215), we define ideology as “a more or less systematic set of ideas that includes the identification of a referent group (a class, ethnic, or other social group), an enunciation of the grievances or challenges that the group confronts, the identification of objectives on behalf of that group (political change—or defense against its threat), and a . . . program of action.” We operationalize this definition by disaggregating ideology along three axes: *conflict framing*, *conception of the ideal polity*, and *territorial aspiration*. Each of these dimensions suggests causal mechanisms that link ideology to rebel alliances.

Conflict framing refers to how rebel factions demarcate the core political, religious, or social categories that constitute one’s in-group and out-groups.² A group’s conflict frame specifies its preferred conflict dyad, the out-group most threatening to the in-group. In Iraq, for example, nationalist insurgents opposed to America’s 2003 invasion of their country employed a resistance frame of Iraqis versus American occupation forces as their primary conflict frame; the Iraqi government and Shiite militias were viewed as mere instruments of America’s occupation. In contrast, jihadist groups, especially Al-Qaeda in Iraq, framed the conflict as a sectarian struggle between Sunnis and Shiites, whereby American forces enabled Shiites to dominate Sunnis. In each instance, conflict framing implies that threats from a particular out-group are more salient than others and that certain parties to the conflict could conceivably cooperate while others are unthinkable; Sunni nationalists could ally with Shiites, whereas sectarian jihadist groups could not. Thus, the conflict frame in-group bounds the choice of allies.

As a group’s conflict frame helps determine whom it attacks, conflict framing may also indirectly promote cooperation to the extent that groups with similar targeting portfolios can more easily cooperate. For example, two rebel groups—one nationalist and one sectarian—may both primarily target the state’s security forces, yet the first casts them as the goons of a tyrannical regime while the second casts those same forces as the soldiers of the rival sect. Although the pair could cooperate on the basis of this common targeting, if the sectarian framing is also extended to justify indiscriminate and controversial attacks against rival sect civilians (included within the nationalist in-group), then the associated dissension would inhibit cooperation.

Conception of the ideal polity is the normatively prescriptive dimension of ideology that orients members to a vision of the desired end state. It specifies how groups define a legitimate sociopolitical order that is worth fighting for, deeming some institutional arrangements appropriate while viewing others as unjust, inequitable, oppressive, or even heretical. This dimension captures the traditional ideological

divides such as the competition between the economic left and right, democrats and authoritarians, and secularists and fundamentalists. When choosing to form alliances, we expect rebels to align with those that offer mutual political corroboration and are working toward similar objectives. Groups with fundamentally divergent postconflict goals or territorial aspirations will have a greater ideological distance to traverse in order to achieve cooperation.

Territorial aspiration delineates the boundaries of the ideal state and orients rebels to the territorial claims of the movement. This dimension captures the degree to which rebels seek to maintain or violate the territorial integrity of their states.³ Movements with shared conceptions of the ideal polity sometimes diverge over the territorial boundaries of that polity. For example, parties representing Basques and Catalans in Spain diverge on the issue of maintaining local autonomy or insisting on separatism as do Scots in the United Kingdom. Arab nationalists in their heyday were divided between advocates of *wataniyya* (homeland patriotism) and *qawmiyya* (pan-Arab unification). Islamists today are divided between those who favor establishing an Islamic order within the framework of the modern national state and those who harbor the irredentist ambition of restoring an Islamic caliphate.

Like the previous dimensions, territorial aspiration is a potential source of unity or division. Separatist groups, for example, may be unwilling to compromise their own territorial demands, creating friction with nationally focused groups. Territorial aspirations are likely to accentuate disagreements between factions as a conflict becomes protracted. Groups that care about the territorial integrity of their states may incline toward a negotiated end to the conflict in order to restore national unity. Those that harbor broader territorial ambitions are less likely to prioritize national unity as the conflict persists and may be inclined to sabotage conflict-ending negotiations.

Agreement on conflict framing, ideal polity, and territorial aspiration, therefore, predict *ideological homophily* in network ties. A fundamental principle of social network analysis, homophily states that “similarity breeds connection,” and social networks tend to be largely homogenous because ties between dissimilar individuals dissolve more quickly (McPherson, Smith-Lovin, and Cook 2001, 415, 16). Homophily prevails because of the presumption of mutual trust and complementarity of interests among actors with uniform attributes (Lichbach 1995, 138-41) and because joining similar others reinforces the cognitive bias toward belief confirmation in polarized political contexts (Balliet et al. 2018). Political homophily has been observed at the individual, organizational, and state levels including life style politics (DellaPosta, Shi, and Macy 2015), online activism (Boutyline and Willer 2017), local government regional planning networks (Gerber, Henry, and Lubell 2013), international trade networks (Maoz 2012), third-party state interventions (Corbetta 2013), and international alliances (Werner and Lemke 1997; Lai and Reiter 2000). We anticipate ideological homophily will also shape rebel alliance choices, yielding our ideology hypothesis:

Hypothesis 1: Ideological alliances: Interrebel cooperation is more likely among ideologically similar groups than ideologically dissimilar ones.

Complementary Logics of Alliance Formation

We consider two logics of alliance composition based on power and state sponsorship, which may operate in conjunction with ideological homophily.

Power in Rebel Alliances

In the most extensive analysis of rebel cooperation, Christia (2012, 240) advances the power-centric theory of MWCs, which are “alliances with enough aggregate power to win the conflict, but with as few partners as possible so that the group can maximize its share of postwar political control.” Absent credible commitments, however, weaker alliance members grow wary of their stronger partner as the alliance nears victory. Hence, the theory predicts coalitional instability as rebels regularly switch sides, thereby maintaining a rough balance of power. Apart from this balancing constraint, the theory remains silent on the composition of the rival coalitions. To the extent that the MWC theory considers the credible commitments problem at its utmost severity, it expects little association between ideology and militant cooperative relationships (Christia 2012, 32, 33).

We propose two contrasting hypotheses about how relative power considerations may affect alliance composition beyond balancing constraints. The first relates to *symmetric alliances* (cooperation between groups of similar power capabilities). Rebel groups in search of greater security may form alliances to aggregate their capabilities against mutual threats.⁴ Given that the pooling of assets and coordination of tactics becomes more difficult as the number of groups grows, two powerful groups can cooperate more efficiently than a coalition consisting of a powerful group and multiple minor groups.⁵ If powerful factions prefer to coordinate with each other, that leaves weak factions to ally with other minor players. Thus, our first power hypothesis predicts:

Hypothesis 2: Symmetric alliances: Interrebel cooperation motivated by mutual security concerns will produce cooperative network ties between groups of comparable power capabilities.

Rebel groups may also seek to maximize their decision-making autonomy in addition to augmenting their capabilities through asymmetric alliances—cooperation between major and minor rebel groups.⁶ Groups that do not feel particularly threatened by the regime may prioritize winning on their own terms. Powerful rebels, in particular, can afford to emphasize enhanced autonomy over security by forming alliances with weaker partners amenable to influence. The weaker faction receives greater security from its alignment with a powerful group, while the dominant rebel faction benefits from both capability aggregation and control over the conduct of minor groups. Thus, we hypothesize:

Hypothesis 3: Asymmetric alliances: Interrebel cooperation motivated by security and autonomy considerations will produce cooperative network ties between groups with dissimilar power capabilities.

State Sponsorship in Rebel Alliances

External sponsorship of rebel movements is a common feature of civil wars.⁷ Rebels covet military, financial, and political support to outmatch the resources of their incumbent regimes, establish international legitimacy, exercise leverage in negotiations, and outcompete rivals. As Gurr (1970, 269) observed long ago, “The greatest potential increment to dissident military capacity is external support.” Indeed, Jones (2017, 136) finds that insurgent movements that receive great power support win nearly half to two-thirds of the time.

External patrons provide arms, money, supplies, or sanctuaries to rebel groups in the expectation that these rebels will exhibit sufficient discipline and solidarity to fulfill their patron’s strategic aims (Salehyan 2010). Bapat and Bond (2012) and Popovic (2018) view external leverage as an important interrebel institution that can help overcome the credible commitments problem, police against side negotiations, and mediate conflicts between rebel groups. This predicts greater interrebel cooperation because sponsors can threaten to withhold financing and war materiel from those who are jeopardizing a cohesive rebel coalition (Lichbach 1995, 179).

However, state sponsors can also undermine rebel unity by incentivizing some rebels to challenge their rivals (Tamm 2016). This is particularly the case when multiple state sponsors with competing political agendas seek to foster their own proxy clients through patronage. The presence of multiple sponsors increases the number of avenues rebel groups have to support themselves and reduces the leverage any individual external patron can exert to foster cohesive rebel coalitions (Salehyan, Siroky, and Wood 2014).

Acknowledging these contradictory effects of state sponsorship on rebel alliances, we propose that two rebel groups that share a single sponsor are more likely to cooperate with one another than dyads with distinct sponsors. This yields our final hypothesis:

Hypothesis 4: State sponsored alliances: Rebel groups that derive support from the same state sponsor will experience greater cooperation than those lacking a common state sponsor.

Table 1 summarizes our hypotheses and suggests their observable implications for the composition of alliance networks at the level of tactical joint operations.

Network Analysis of Syrian Militant Alliances

We employ social network analysis to test our four theoretical propositions. A social network consists of nodes and the ties between node dyads. The nodes can represent

Table 1. Hypotheses, Causal Mechanisms, and Expected Network Structure.

Hypotheses	Causal Mechanisms	Expected Network Outcomes
Hypothesis 1: Ideological alliances	<i>Ideological homophily</i> shapes cooperation due to similar understanding of enemies and allies (conflict framing), ideas and institutions of sociopolitical order (ideal polity), and the boundaries of that order (territorial aspiration)	Network structure will be shaped by groups' shared ideological attributes
Hypothesis 2: Symmetric alliances	<i>Power aggregation</i> is the primary consideration behind cooperation	Network structure will be shaped by groups' comparable power
Hypothesis 3: Asymmetric alliances	<i>Security-autonomy trade-off</i> is the primary consideration behind cooperation	Network structure will be shaped by groups' disparate power
Hypothesis 4: State sponsored alliances	<i>Sharing state sponsors</i> compels rebels to forge cohesive alliances	Network structure will be shaped by groups' shared state sponsors

individuals, organizations, or states and ties can correspond to relationships such as communication, cooperation, and conflict. Social network analysis can account for the interdependence of relationships within a set of political actors (Ward, Stovel, and Sacks 2011; Hafner-Burton, Kahler, and Montgomery 2009). Alliance models that assume independence of observations in dyads miss out on relational data because alliances are not created in a vacuum; they are dependent upon relationships with multiple groups (Dorussen, Gartzke, and Westerwinter 2016).

The fragmented nature of civil conflicts implies network analysis should be a fruitful methodology for addressing militant behaviors such as alliance formation and infighting (Zech and Gabbay 2016). We illustrate the utility of the network approach in our analysis of rebel alliances in the Syrian civil war.

Rebel Factions in the Syrian Civil War

In March 2011, Arab Spring protest waves reached Syria after making their way from North Africa to the Middle East. Bashar al-Assad's regime initially sought to quell protests and prevent their diffusion through a mix of repression and concessions. However, these measures failed as protests gained momentum across Syria's major cities, and the protestors' demands shifted from reforms to regime change. As the conflict became militarized, the Free Syrian Army (FSA) formed from the ranks of defecting officers and its affiliated brigades began engaging in conventional armed attacks against regime forces. The FSA exemplified the secular nationalist tendency, framing the Syrian rebellion as a

national and democratic revolution that encompasses Syria's diverse ethnic and religious communities.

The inability of protesters and the FSA to topple the Assad regime in the opening months of the insurgency gave rise to rival armed factions, the most notable of which was the Al-Qaeda affiliated Al-Nusrah Front (ANF). Formed in January 2012, ANF was avowedly sectarian and jihadist, casting the conflict not as a revolution, but rather as a holy war against a secular regime dominated by heretical Alawites (an offshoot sect of Shiism). It called for the formation of an Islamic state strictly adherent to religious law (International Crisis Group 2012; Lister 2015).

Many other Islamist factions emerged, ranging from Muslim Brotherhood sympathizers such as Al-Tawhid Brigade to Salafists such as Ahrar al-Sham Islamic Movement (ASIM). The latter became one of the dominant factions in the insurgency, competing with both the FSA and ANF (International Crisis Group 2012). ASIM represented Salafist nationalists that wanted to establish an Islamic state within the boundaries of Syria's national territory, but, unlike ANF, it did not frame the conflict in sectarian terms.

Kurdish communities established their own armed groups, notably the People's Protection Units (Yekîneyên Parastina Gel, YPG), to defend their territories from regime forces as well as hostile rebels (International Crisis Group 2014). The YPG is secular in orientation and views Kurdish co-ethnics as its primary in-group for whom it seeks autonomy within, or secession from, the Syrian state.

In 2013, the Islamic State in Iraq and the Levant (ISIL), led by Abu Bakr al-Baghdadi, splintered the ranks of ANF to form an even more extreme sectarian jihadist faction. ISIL further aggravated the conflict by intensifying sectarian polarization, expanding the conflict into neighboring states, and threatening international security through global terrorism. ISIL also produced fratricidal violence within the rebel movement as it sought to expel rivals from its strongholds and asserted itself as the sole legitimate rebel organization that merits allegiance (Lister 2015).

Cooperative Rebel Network: Data and Variables

We measure rebel cooperation (our dependent variable) in terms of claims of tactical joint operations. The use of joint operations provides a more demanding test of ideological homophily than formal alliances because if groups prefer to cooperate with ideologically similar rebels at the tactical level, then they should be even more selective for the deeper, leadership-level collaboration required in strategic coalitions.

We began our data collection by tracking the operational claims of forty-four major rebel groups using Arabic and English newspapers of record, US government informational briefs, and think tank reports. Since it was not possible to collect data on all the Syrian rebel groups, we limited our analysis to a medium *N* that had sufficient credible information to ensure data reliability. Although not ideal, expanding the analysis to less prominent groups risked sacrificing quality for quantity.

Furthermore, by focusing on the primary rebel actors, we assumed as Krause (2017, 14) does that prominent players matter the most and that minor players are less likely to shape conflict trajectories.

We used automated text processing to find claims that contained “joint,” “collaboration,” “cooperation,” or “support” and then hand coded each claim to verify it constituted a joint operation. We collected their claims of attacks—including both targets of attack and groups involved in joint operations. These data come from US Government translations of insurgents’ statements and operational claims, drawn from social media (Facebook, Twitter, and YouTube) as well as various jihadist forums.

We used any claims of joint operations from the 44 organizations to one another or to smaller groups to generate a network of some 220 Sunni Arab and Kurdish groups actively engaged in the conflict. This resulted in 696 joint operations and more than 930 ties between the 220 groups across the four years of the conflict spanned by our data (July 2012 to June 2015). The joint operations network is a symmetric matrix whose elements are the total number of joint operations claimed by either group in the dyad represented. If more than two groups were claimed to be involved in an attack, we gave each group a tie with each other group mentioned.

Network Description

The full network (Figure 1) shows some clear patterns of cooperation. The more prominent groups, such as ASIM, ANF, ISIL, and Al-Sham Legion (ASL), have separate retinues of small groups linked only to them. However, there is also cooperation among prominent groups. ASIM, the group with the most ties, cooperates with large FSA-affiliated groups such as the Ahfad al-Rasul Brigades (AARB) and the Al-Furqan Battalions (AFB) as well as the sectarian jihadists ANF and ISIL. There are Kurdish groups in our data, observed in the lower left corner and linked to the Sunni Arab militants by a single connection—YPG to Ar-Raqqa Revolutionaries Brigade (ARRB).

Measuring Ideology, Power, and State Sponsorship in the Network

We evaluated rebel groups for three ideological areas of relevance to the Syrian conflict. Sectarianism serves as our conflict frame variable: groups with high sectarianism scores cast the conflict as Sunnis versus Shiites/Alawites, whereas groups with low sectarianism scores have little or no anti-Shiite rhetoric. Salafism, which measures the extent to which groups ascribe to that puritanical strain of Sunni Islam, provides our ideal polity variable. The use of Salafism better resolves differences within various stripes of Islamists than a simple secularism versus Islamism scale. Revisionism is used for the territorial aspiration component of ideology: groups with low scores seek to preserve Syria’s territorial integrity, whereas a high score

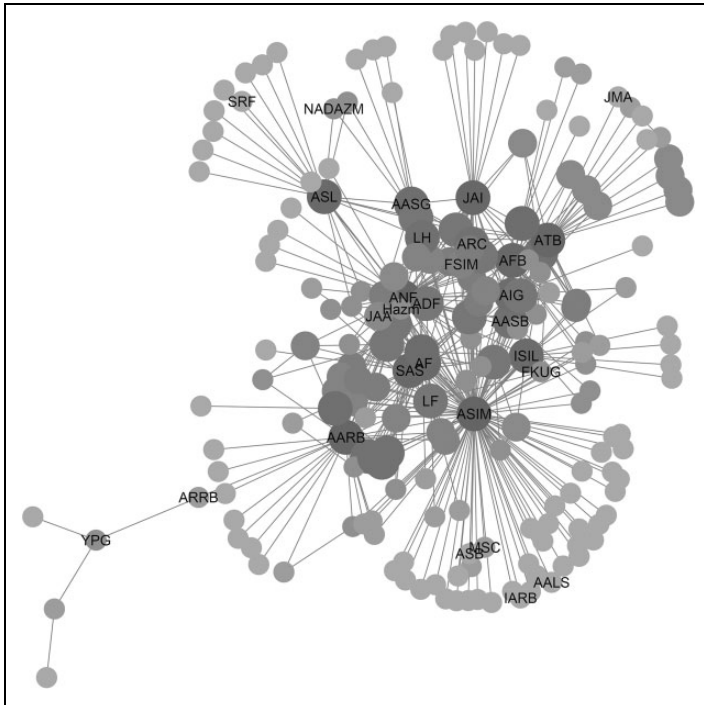


Figure 1. Network of Syrian militant joint operations. Links indicate presence of one or more ties between groups. Circle size is proportional to node degree. The names of smaller groups have been removed from this graph to make it readable.

signifies a desire to abrogate it, in particular as do Caliphate-minded sectarian jihadists or Kurdish separatists.

Each axis of ideology is coded on a scale of 1 to 5, a range that allows us to capture the proximity or distance of groups on each component. We hand coded the ideology of the forty-four Syrian rebel groups using manual coding from the groups' founding charters and other public declarations (see Supplemental Material for coding methodology and the rebel groups' ideological scores). We aggregate the scores of the three components of ideology into one average ideology score and check those results to make sure the variable we have constructed makes sense in light of Syria's factional divides (see Supplemental Material).

Our methodology situates groups in ways that make sense in the context of the Syrian civil war. We would expect groups in Syria to broadly fall into the following categories: *Secular nationalists*, *Salafist nationalists*, *secular Kurdish separatists*, and *sectarian jihadists* (see Table 2).⁸

Table 2. Ideological Spectrum of Syria's Militant Factions.

Dimensions of Ideology	Secular Nationalist	Kurdish Separatist	Salafist Nationalist	Sectarian Jihadist
Conflict frame	Syrians versus the Assad regime	Kurds vs. sectarian jihadists	Sunni Syrians vs. the Assad regime	Sunnis vs. Alawites/Shiites
Ideal polity	Secular democratic Syria	Kurdish secular government	Islamic state	Islamic state
Territorial aspiration	Unified Syria	Separate republic or autonomy	Unified Syria	Transnational Islamic caliphate

We measure power in the network by group size. For each group, we collected as many estimates of size as were available (see Supplemental Material). We created a low–medium–high estimate for each group, when possible, and used the medium estimate in our analysis. As we could not locate size estimates for a few small groups, they were assigned a minimal value of 500 fighters.

Although group size is by no means a comprehensive measure of power, it is often used as such in statistical analysis (Akcinaroglu 2012; Christia 2012; Krause 2013/2014). We make the assumption that groups that can mobilize more fighters than their competitors are also likely to have substantial financial resources to arm those fighters, pay them salaries, and support their families. Thus, we proceed with group size as a proxy for other elements of rebel power. We also use the Institute for the Study of War (ISW) “powerbrokers” measure to validate this variable and find that no group coded as a regional powerbroker by ISW is also a “small” rebel group in terms of number of fighters (Cafarella and Casagrande 2016). The smallest group in our data listed as a powerbroker is Nur al-Din al-Zinki Movement with a size estimate of approximately 5,000 fighters.

Lastly, we assess the presence or absence of shared state sponsorship by drawing upon informed secondary sources that identify the primary sponsors of rebel groups (see Supplemental Material for the complete list of sources). As for rebel group location, we used the operational claims of rebels to determine their primary areas of operation. Some groups operated locally, and were coded as such, while others had multiple branches. Groups that appeared in four or more governorates were coded as national, even though they may not have had presence in every Syrian region.

The Core Network

We coded covariate data for forty-four rebel groups and tracked the collaborative relationships among them. Only thirty-one of those forty-four groups participated in collaborative tactical relationships, so we proceed with 376 ties among these thirty-one groups (see Figure 2).

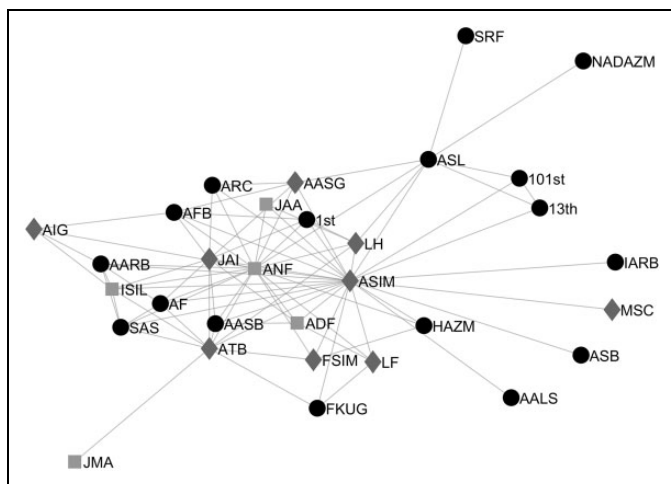


Figure 2. Diagram of the thirty-one core groups with collaborative ties within the network. A line between two groups indicates the presence of at least one joint operation between them. Node shapes denote secular nationalist (circles), Salafist nationalist (diamonds), and sectarian jihadist (squares) ideological classifications.

The Supplemental Material provides data on the core groups used in this analysis, as well as descriptive statistics. The core groups display significant variation in terms of group size, ideology, and state sponsorship.

Network Regression Analysis

To evaluate our hypotheses relative to one another and while controlling for additional variables, we run an AME regression models from Minhas, Hoff, and Ward (2019). The AME regression is an extension of the class of network inference methods known as latent space models, which seek to relate the tendency of nodes to form ties with each other to their proximity in an underlying space of latent variables (Cranmer et al. 2017). Rather than assuming independence of observations (as per de Finetti's theorem, which justifies the conditionally independent and identically distributed assumption in statistics), these models account for dependence between dyads (row and column means—additive effects) and higher-order dependence in the network structure such as stochastic equivalence (multiplicative effects—see Supplemental Material for model specifications and details; also see Hoff 2015). AME models have been used recently to analyze conflict and international relations data (see Dorff 2015; Dorff, Gallop, and Minhas Forthcoming; Minhas, Hoff, and Ward 2016).

Table 3 presents the AME regression results for difference in average ideology, difference in power, and shared state sponsorship separately (models 1–3) and

Table 3. Square Root Transformed Dependent Variable.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	.07 (.11)	.75 (.00)	−.00 (.00)	−.48 (.28)	−.41 (.32)
State sponsorship (node)				.07 (.11)	.09 (.12)
Ave. ideology (node)				.07 (.04)	.07 (.04)
Power (node)				.01 (.01)	.00 (.01)
ASIM (node)					.32 (.28)
Ideol. diff. (dyad)	−.04*** (.01)			−.05*** (.00)	−.05*** (.01)
Power diff. (dyad)		−.06* (.04)		−.01*** (.00)	−.01*** (.00)
Shared St. sponsor (dyad)			.06 (.05)	−.00 (.05)	−.10 (.05)
Shared location (dyad)				.17*** (.04)	.17*** (.04)

Note: Results of additive and multiplicative effects regression analysis. Dependent variable is square root of the count of collaborative ties. Standard errors are given in parenthesis.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

together with controls for their node-level values (for state sponsorship, a dummy variable corresponding to the presence or absence of a sponsor) and shared location (model 4), and a control for our most prominent group, ASIM (model 5). As the AME software does not yet cover Poisson or Negative Binomial Distributions, we follow standard practice for such cases and take the square root of the dependent variable to account for the progressively increasing residuals in our data as independent variable values increase. Results using the direct count value of the dependent variable and an ordinal model are displayed in Supplemental Material.

Table 3 reveals strong support for the ideological alliances hypothesis (Hypothesis 1): groups that are ideologically proximate cooperate with each other more so than the ones that are ideologically distant. The decrease in cooperation with increasing ideological difference is statistically significant regardless of the inclusion of other covariates and whether the square root transformation or raw counts is used.

We find some support for power symmetry (Hypothesis 2): groups of similar strength tend to cooperate with each other more so than those that vary in their power capabilities. This finding, however, is inconsistent. It is always significant in the square root transformation but, in the direct tie count, only becomes significant for model 4 (see Supplemental Material). The support for power symmetry rules out the opposite prediction of power asymmetry (Hypothesis 3).

Ideology has a consistent, statistically significant value across models with a smaller degree of uncertainty than for the power difference (see Supplemental Material) and with a larger effect size than power in the majority of models. Also, the robustness checks below firmly support the ideological homophily hypothesis, but not power symmetry or asymmetry.

We do not find evidence in favor of the state-sponsored alliances hypothesis (Hypothesis 4), which predicts that rebel dyads with shared sponsors should exhibit greater cooperation than dyads with distinct external sponsors. Perhaps this result is an artifact of the lack of weight in shared sponsorship; our data represent whether two groups were ever sponsored by the same state, with no weight given to how important a sponsor was for a particular group. More substantively, however, two plausible explanations may shed light on this finding. First, as long as rebel groups were generally cooperating with ideologically similar groups, which is what we find in the Syrian civil war, their state sponsors may not have cared if their clients were working with rebels that have different sponsors. However, had it been the case that Syrian groups were, generally speaking, cooperating with ideologically dissimilar factions, their state sponsors would have exerted pressure to break those cooperative ties.

A second possible explanation is that tactical cooperation in the form of joint operations is less visible to state sponsors than strategic mergers or formal coalitions. States may have overlooked their clients' tactical cooperation partners in the Syrian theater to achieve the broader objective of regime change. It is more likely that overlapping external sponsorship plays a greater role in facilitating or hindering strategic alliances that are much more formal and public than they do tactical cooperation. Therefore, our finding regarding state sponsorship at the tactical cooperation level does not preclude the importance of this variable in strategic alliances, which we do not explore in this study.⁹

Table 3 includes shared location as a control since it is possible that observed cooperation between ideologically similar groups is merely a surface manifestation of the underlying ideological homogeneity of groups who operate in the same area. Tactical joint operations, perforce, require rebels to operate in the same location, and if these operating areas consisted only of groups with similar ideologies, then ideological homophily would arise simply due to geographic proximity. One might argue that the homogeneity of a given geographic area with respect to its ethnic or religious composition would foster such ideological homogeneity. Alternatively, one might expect that social influence between proximate groups would result in ideological convergence. This argument, however, begs the question as to the epiphenomenal nature of ideology by assuming that it is easily malleable in the first place. In counterpoint, contact between ideologically distinct groups may readily result in their violent conflict rather than ideological convergence, an outcome amply demonstrated in the Syrian conflict.

Empirically, Table 3 supports shared location as being important to tactical collaboration as intuitively expected, but ideological homophily still remains significant. The above supposition that geographic proximity imposes ideological uniformity is at odds with the fact that the predominantly Sunni Arab composition of the rebel movement reflected ideological diversity—a diversity that existed in close geographic proximity. In Aleppo, for example, there were no less than twenty-two

separate armed groups representing three distinct ideological strands: secular nationalism, Salafist nationalism, and sectarian jihadism. Similarly, in Idlib, there were sixteen groups representing these three distinct ideologies. In fact, nine of the ten Syrian governorates in our study had at least two ideologically divergent groups (see Supplemental Material).

Robustness Checks: Assortativity and Network Simulation

Two additional network methods are employed as robustness tests for ideology and power in the above analysis: (1) comparing the assortativity, a metric of network homophily or heterophily, of the observed network with the distribution obtained from a null model simulation and (2) a simulation with homophily included to estimate the characteristic ideological or power scale over which cooperation is more likely (a heterophily simulation is used if power asymmetry is indicated). These two methods consider the three ideological components, average ideology, and power—all treated separately.

Assortativity is the standard metric for assessing whether tie formation is driven by similarity or dissimilarity with respect to a scalar variable (as we operationalize power and ideology). The assortativity α is the correlation of the variable values for the nodes at each end of a tie taken over all ties (see Network Simulation Appendix in the Supplemental Material). An α value of +1 signifies maximal homophily whereas -1 represents maximal heterophily. For statistical testing purposes, the assortativity cannot be treated as one would treat a standard correlation because ties are not taken to be independent.

We developed a simple simulation of the tie formation process that can be implemented using our empirical data. When ideology (or power) is not included, the simulation acts as a null model that can generate a distribution of assortativity values for calculating the statistical significance of the observed assortativity. Our simulation-based tests will decide that homophily (heterophily) is present when the difference between the empirical network assortativity and the mean of the null simulation over many runs is positive (negative) and statistically significant.

In the simulation, nodes form ties (i.e., groups conduct joint operations) probabilistically. Each iteration consists of the placement of a tie between nodes where the iterations proceed up to the total number of ties in the observed network. The simulation is constrained in that it seeks to reproduce the node degrees in the observed network. Each node can only receive a maximum number of ties equal to its observed degree (its number of joint operations).¹⁰ The model essentially assumes that a group wishes to make its units available for a certain number of joint operations over a given time period. The more units available at a given moment, the more likely a group is to find a partner, which, at the dyad level, implies that the interaction probability between a pair of groups depends on the product of their available units.

Table 4. Robustness Analysis Results.

Variable	Assortativity				Simulation	
	α	α_{null}	σ_{null}	p	l/l_s	CI
2012–2015 ($N = 376$)						
Conflict frame	-.032(+)**	-.196	.039	<.0001	2.3	[2.0, 2.7]
Ideal polity	-.096(+)	-.097	.047	.97	—	—
Territorial aspiration	-.027(+)**	-.238	.042	<.0001	2.3	[2.0, 2.8]
Average ideology	.017(+)**	-.145	.042	<.0001	2.3	[1.7, 3.6]
Power	-.235(-)	-.172	.048	.19	—	—
2012 ($N = 55$)						
Average ideology	-.138 (-)	-.134	.108	.96	3.1	[1.9, 5.4]
Power	-.574 (-)	-.382	.110	.07	5,900 (l_s)	[2,500, 10,000]
2013 ($N = 136$)						
Average ideology	.096 (+)**	-.122	.072	.003	1.8	[1.5, 2.2]
Power	-.31 (-)*	-.152	.078	.04	6,300 (l_s)	[3,000, 11,000]
2014 ($N = 119$)						
Average ideology	.025 (+)**	-.230	.075	.0006	—	—
Power	-.236 (-)	-.176	.081	.47	—	—
2015 ($N = 66$)						
Average ideology	-.313 (-)	-.242	.104	.51	—	—
Power	-.141 (+)	-.146	.107	.94	8,900	[6,800, 14,000]

Note: N is the number of ties. For assortativity, α is the assortativity of the observed network where + (–) indicates α greater (less) than α_{null} corresponding to homophily (heterophily); α_{null} and σ_{null} are, respectively, the mean and standard deviation of the assortativity in the null simulation taken over 10,000 runs; the p value is the (two-tailed) fraction of runs exceeding $|\alpha - \alpha_{\text{null}}|$. For Simulation, l is the mean interaction length (suppression length l_s where indicated) and CI is the 95 percent confidence interval (blank entries signify the absence of a clear minimum); 1,000 runs at each point were used to generate 1,000 resamples of size 50 with replacement and then the minimum l (or l_s) for each resample was found.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

There are three variants of the simulation (see Network Simulation Appendix in the Supplemental Material): (1) a null simulation in which, as described above, only the node degrees affect tie formation, not the node variables, (2) a homophily simulation in which the probability of tie formation between two nodes decreases as the distance between their variable (ideology or power) values gets larger, and (3) a heterophily simulation in which the probability of tie formation increases with increasing distance between them. For the homophily simulation, a parameter called the interaction length, l , sets the characteristic length scale so that, roughly speaking, nodes are significantly more likely to form ties within that length from each other than beyond that range. As the interaction length scale increases, the effect of homophily diminishes until the null model is effectively recovered. The heterophily simulation uses a different parameter, the suppression length l_s , for which the

probability of interaction is reduced for the region within the suppression length and much greater in the region outside it.

Robustness Findings

The statistical tests using assortativity are shown in Table 4 for the entire 2012 to 2015 period and by individual years (note that 2012 and 2015 are not full years of data). For the entire period, the observed assortativity values for ideology are all greater than the mean of the null distribution indicating homophily. Conflict frame and territorial aspiration are highly significant. Although ideal polity is not significant, the mean of all three components, average ideology, remains highly significant. Accordingly, the assortativity tests support the ideological homophily hypothesis (Hypothesis 1); ideological clustering characterizes the network structure.¹¹ The results of the simulation, which models interactions driven by ideological homophily, are similar in that conflict frame, territorial aspiration, and average ideology all display well-defined values of the interaction length l whose confidence intervals are less than the full range of the ideology scale, whereas ideal polity does not. The common value of $l = 2.3$ indicates that the zone in which cooperation is relatively likely is about half the length of the full ideology scale. Thus, groups in the middle of the spectrum can cooperate with both ends, but cooperation between the opposed extremes of the scale will be much less common.

For the network by year, the assortativity for average ideology is highly significant for 2013 and 2014, the two years with the greatest number of ties. The 2012 and 2015 networks, which are smaller, show no significance.

The substantive effect of ideological homophily can be assessed by running the homophily simulation with the estimated interaction length from Table 4 and examining how the number of ties for a dyad depends upon their ideological separation. For example, the interaction length for average ideology in the 2012 to 2015 network is 2.3. To enable a more generic assessment not contingent upon the specific Syria configuration, we simulated a network with a uniform degree distribution and uniform ideology distribution and found that increasing the distance between two nodes initially collocated at the middle of the ideology range to successive distances of (1, 2, 3, 4) units decreased the probability of tie formation between them by (5.4, 20.5, 39.2, 56.6) percent relative to the probability at zero distance (see Supplemental Material). Although the probability of tie formation depends nonlinearly on the ideological distance, averaging the above changes yields a 14.15 percent drop in probability per unit of distance. This value is consistent with the 13.4 percent decrease per one unit shift in ideological difference found by taking the β value from AME model 5 and running a simple linear prediction function while varying the values of ideological difference.

Turning to power, the observed assortativity for the full period is less than the mean of the null simulation, indicating a tendency toward heterophily, but is not significant. Given this tendency, the heterophily simulation was performed, but no

well-defined suppression length was found. The assortativity tests for the first three individual years also indicate a heterophily tendency, but which only rises to significance for 2013. On balance, the assortativity tests do not support either power symmetry or asymmetry.

When comparing the findings of the AME and robustness analyses, both show statistically significant results for ideological homophily, thus providing comprehensive support for ideology as an important factor in determining rebel collaboration. However, the two analyses differ over power. The AME analysis indicates statistically significant power symmetry while the assortativity analysis points toward power asymmetry, albeit not significant. The reason for this disparity may arise from a nonlinear relationship between tie formation and group power differentials. Considering the difference between the observed number of ties in a dyad and that expected based solely from the group degrees (Equation 1 in the Supplemental Material), the distribution of this quantity as calculated from the network (no simulation involved) shows an inverted U-shape as a function of power difference rather than a monotonic decrease (increase) as would be fully consistent with a power symmetry (asymmetry) dynamic (see Supplemental Material): the observed-expected ties difference is negative at both low ($<5,000$) and high ($>20,000$) power differences and positive for intermediate ones. The tie suppression at low power difference is consistent with power asymmetry whereas that at high power difference is consistent with power symmetry. In contrast, a similar plot for ideology shows a greater than expected number of ties for low ideology differences (<2) and smaller than expected for higher ideology differences and so is clearly consistent with an overall homophily effect.

Finally, we address the concern that the finding of ideological cooperation may be an artifact of our limited sample of only forty-four groups of hundreds in the Syrian conflict. As the set of omitted groups is almost entirely, if not completely, composed of relatively weak groups, this concern amounts to the possibility that the weaker groups in our sample are unrepresentative of the broader universe of weak groups in Syria. Since the strong groups are ultimately of greatest importance, we test for ideological homophily between them. Indeed, considering the network of joint operations between groups of size at least 5,000 over the full time period, the assortativity test finds homophily for average ideology to be significant ($p < .0001$). Therefore, ideological similarity helps drive cooperation between the groups whose behavior is most consequential to the conflict. Additionally, a t test reveals no significant difference between the average ideology means of the strong and weak ($<5,000$) groups in our sample, so there is no basis to believe that our sample of weak groups is unrepresentative.

Discussion and Conclusion

Rebel cooperation is a common occurrence in civil conflicts. In this study, we wanted to know with whom rebels cooperate in the context of fragmented conflicts

that feature a diversity of ideological actors, variation in group-level power capabilities, and a plethora of state sponsors. Theoretically, we proposed three components of militant ideology and argued how each can facilitate cooperation. Conflict framing promotes shared understandings about in-groups and out-groups, thereby easing potential dissension about permissible allies and targets. A similar conception of the ideal polity encourages groups to work cooperatively toward achieving compatible visions of the postconflict political order. Territorial aspiration impacts fundamental questions such as whether or not rebels seek to break up the national state or maintain its integrity, which are incompatible goals that dampen cooperation. We employed an innovative network-analytic methodology, constructing a militant tactical cooperation network from claims of joint operations and relating its structure to ideology, power, and state sponsorship in the Syrian civil war.

Substantively, our central finding is that ideology is an important determinant of alliance composition in the Syrian civil war. Groups that were ideologically similar cooperated more frequently than those who were ideologically dissimilar: according to our models, a one unit increase in ideological distance corresponds to about a 14 percent decrease in the likelihood of rebel tactical cooperation. Syrian groups in the middle of the ideological spectrum were willing to cooperate with groups at the end of the spectrum, but groups at the end of the ideological spectrum were less willing to cooperate with each other. No clear finding concerning power emerged as one analysis supported power symmetry while the other supported neither symmetry or asymmetry. We found no evidence that having a common state sponsor encouraged cooperation.

Our ideology hypothesis and results may elicit an endogeneity objection. It could be asserted that stable interrebel relationships motivated by power form first and then groups adjust their ideologies accordingly. In this scenario, power drives the ideological preferences exhibited in alliance composition. This challenge assumes that militant groups arise as ideological blank slates, contrary to the fact that the founders of such groups often have strong ideological orientations. Many of the individuals who formed Syria's major Islamist rebel groups were actually in jail at the start of the revolution due to their prior Islamist activism and then subsequently released (Lister 2015, 53-55; Baczko, Dorronsoro, and Quesnay 2018, 184).¹² In addition, ideological charters, an important element of our coding, are typically issued by groups shortly after their formation. Their ideological statements, therefore, are biased toward a time before these groups have formed cooperative ties with other rebels, and so evidence of homophily in the network reflects selection of similar others.

Another endogeneity concern is that we treat ideologies as fixed, but conflict processes are likely to change the ideological preferences of rebel factions over time. We treat the question of ideological change as an empirical one and our operationalization of ideology using the three components can help track that change. We suspect that ideologies change over time but do so gradually. A common process in social networks is increasing homogeneity in network ties because of the selection of similar partners and the reinforcing effect of social influence of those partners in

maintaining that similarity. Ideological shifts, therefore, will typically be limited and evolutionary, a process that allows for stable patterns to emerge between ideology and cooperative network structure. This view of incremental ideological change is supported by our findings of ideological homophily for the full 2012 to 2015 network and individually for 2013 and 2014 using ideology scores biased toward earlier times in group histories.

Another concern relates to the interaction between power balancing and ideological considerations. These broad concepts are not alternative and incompatible explanations of alliance composition. Rebels may form balanced coalitions, each of which consists of ideologically similar groups. As rebels face a greater (lesser) threat from the state, they may become less (more) ideologically selective about their allies. However, ideology may also act as a barrier to alliance formation even when the distribution of power is so adverse that it would seem to demand rebel unification. In Syria, the tide turned dramatically against the rebels after Russia's direct military involvement on the side of the regime in late 2015 and after the fall of Aleppo in late 2016. Yet the rebels did not ally across ideological lines but remained bitterly divided (Collins 2017; Perry and Al-Khalidi 2017).

Our empirical analysis of a single case study limits our ability to generalize beyond the Syrian conflict. Although not entirely unique, the Syrian civil war is characterized by severe levels of movement fragmentation, a wide spectrum of ideologies, and a perplexing array of external interventions by state and substate actors. Therefore, it may not be entirely representative of other civil wars where rebel groups are fewer in number, nonideological identities prevail (such as in ethnic or resource-based conflicts), or where international interference is limited in scope. Our findings regarding the robustness of ideological homophily in Syrian militant alliances should be thoroughly investigated in other conflicts to have confidence in its generalizability. Our Syrian study, however, highlights the need to consider seriously the role of ideology in rebel alliances and offers a template for researching civil conflicts that exhibit similar patterns of intense fragmentation, ideological polarization, and tactical alliances such as those ongoing in Ukraine, Iraq, Afghanistan, Pakistan, Libya, Sudan, and Yemen.

Acknowledgments

We thank the two anonymous reviewers. We also thank Cassy Dorff, Peter Hoff, Peter Krause, Lee Seymour, Sarah Parkinson, Jon Mercer, Theo MacLauchlan, and Mike Weintraub for their helpful comments.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Defense Threat Reduction Agency and the Office of Naval Research under grants HDTRA1-10-1-0075, N00014-15-1-2549, and N00014-16-1-2919.

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Ideological diversity emerges for a number of reasons. First, some conflicts are ideological at their core, such as the ones featuring communist and fascists in the twentieth century (e.g., Spanish civil war 1936–1939), or fundamentalists and secularists today. Second, prewar political mobilization based on ideological cleavages can extend into the civil war, shaping dynamics of rebel cohesion (Staniland 2014) and violence between ideological rivals (Balcells 2017). Third, the entry of transnational ideological actors such as al-Qaeda or Hezbollah can create ideological polarization, forcing rebel groups and their communities to take sides based on sharp ideological divides (Bakke 2014).
2. Fearon and Laitin (2000, 857) refer to this process as constructing “antagonistic identities,” Asal and Karl Rethemeyer (2008, 438) refer to it as “othering,” and Shesterinina (2016, 417) terms it “collective threat framing.”
3. Territorial aspirations have been at the root of many secessionist civil conflicts, resulting in 131 sovereign states coming into existence since 1945, “a threefold increase in 70 years” (Griffiths 2016, 1).
4. Between states, Morrow (1991, 921-23) argues that capability aggregation drives symmetric alliances.
5. This assumption is supported by Lichbach’s (1995, 19) observation that in rebel coalitions, the largest and richest organizations tend to pay a disproportionate cost for maintaining an alliance.
6. In international relations, Morrow (1991, 921-23) finds alliances are more frequent between powerful and weak states than between those of comparable power.
7. According to Jones (2017, 136), of 181 insurgencies between 1946 and 2015, 82 percent involved outside support.
8. For similar categorization of rebel factions, see Cafarella and Casagrande (2016, 9) and Phillips (2016, 131-34).
9. Additional model specifications, including bivariate relationships and additive and multiplicative effects diagnostic plots, are available in the Supplemental Material.
10. It is not always possible to reproduce the degrees exactly, but the differences are typically small.
11. A network visualization showing how groups cluster by ideology is included in the Supplemental Material.

12. Ahrar al-Sham Islamic Movement and Jaysh al-Islam are two such groups. It is likely that the Assad regime released these leaders in a cynical ploy to affirm its narrative that the opposition consisted of jihadist terrorists, a strategy that implies that the regime, at least, believed that these men would act on their ideological predilections. Moreover, veterans of earlier jihads formed two other prominent factions, Al-Nusrah Front and the Islamic State, which suggests deep ideological commitments over time.

References

- Akcinaroglu, Seden. 2012. "Rebel Interdependencies and Civil War Outcomes." *Journal of Conflict Resolution* 56 (5): 879-903.
- Asal, Victor, and R. Karl Rethemeyer. 2008. "The Nature of the Beast: Organizational Structures and the Lethality of Terrorist Attacks." *Journal of Politics* 70 (2): 437-49.
- Baczko, Adam, Gilles Dorronsoro, and Arthur Quesnay. 2018. *Civil War in Syria: Mobilization and Competing Social Orders*. New York: Cambridge University Press.
- Bakke, Kristin M. 2014. "Help Wanted? The Mixed Record of Foreign Fighters in Domestic Insurgencies." *International Security* 38 (4): 150-87.
- Bakke, Kristin M., Kathleen Gallagher Cunningham, and Lee J. M. Seymour. 2012. "A Plague of Initials: Fragmentation, Cohesion, and Infighting in Civil Wars." *Perspectives on Politics* 10 (2): 265-83.
- Balcells, Laia. 2017. *Rivalry and Revenge: The Politics of Violence during Civil War*. New York: Cambridge University Press.
- Balliet, Daniel, Joshua M. Tybur, Junhui Wu, Christian Antonellis, and Paul A. M. Van Lange. 2018. "Political Ideology, Trust, and Cooperation: In-group Favoritism among Republicans and Democrats during a US National Election." *Journal of Conflict Resolution* 62(4): 797-818.
- Bapat, Navin A., and Kanisha D. Bond. 2012. "Alliances between Militant Groups." *British Journal of Political Science* 42 (4): 793-824.
- Boutyline, Andrei, and Robb Willer. 2017. "The Social Structure of Political Echo Chambers: Variation in Ideological Homophily in Online Networks." *Political Psychology* 38 (3): 551-69.
- Cafarella, Jennifer, and Genevieve Casagrande. 2016. *Syrian Armed Opposition Power-brokers*. Accessed December 2, 2017. http://www.understandingwar.org/sites/default/files/Syrian%20Armed%20Opposition%20Powerbrokers_0_0.pdf.
- Christia, Fotini. 2012. *Alliance Formation in Civil Wars*. New York: Cambridge University Press.
- Collins, Dylan. 2017. "Idlib's Rebel Split: A Crossroads for Syrian Opposition." Accessed May 14, 2018. <https://www.aljazeera.com/indepth/features/2017/01/idlib-rebel-split-crossroads-syrian-opposition-170131134053716.html>.
- Corbetta, Renato. 2013. "Cooperative and Antagonistic Networks: Multidimensional Affinity and Intervention in Ongoing Conflicts, 1946-2001." *International Studies Quarterly* 57 (2): 370-84.
- Costalli, Stefano, and Andrea Ruggeri. 2015. "Indignation, Ideologies, and Armed Mobilization: Civil War in Italy, 1943-45." *International Security* 40 (2): 119-57.

- Cranmer, Skyler J., Philip Leifeld, Scott D. McClurg, and Meredith Rolfe. 2017. "Navigating the Range of Statistical Tools for Inferential Network Analysis." *American Journal of Political Science* 61 (1): 237-51.
- Cunningham, Kathleen Gallagher. 2013. "Actor Fragmentation and Civil War Bargaining: How Internal Divisions Generate Civil Conflict." *American Journal of Political Science* 57 (3): 659-72.
- Cunningham, Kathleen Gallagher, Kristin M. Bakke, and Lee J. M. Seymour. 2012. "Shirts Today, Skins Tomorrow: Dual Contests and the Effects of Fragmentation in Self-determination." *Journal of Conflict Resolution* 56 (1): 67-93.
- DellaPosta, Daniel, Yongren Shi, and Michael Macy. 2015. "Why Do Liberals Drink Lattes?" *The American Journal of Sociology* 120 (5): 1473-511.
- Dorff, Cassy L. 2015. "Civilian Autonomy and Resilience in the Midst of Armed Conflict." PhD Dissertation. Duke University.
- Dorff, Cassy L., Max Gallop, and Shahryar Minhas. Forthcoming. "Networks of Violence: Predicting Conflict in Nigeria." *Journal of Politics*.
- Dorussen, Han, Erik A. Gartzke, and Oliver Westerwinter. 2016. "Networked International Politics: Complex Interdependence and the Diffusion of Conflict and Peace." *Journal of Peace Research* 53 (3): 283-91.
- Driscoll, Jesse. 2012. "Commitment Problems or Bidding Wars? Rebel Fragmentation as Peace Building." *Journal of Conflict Resolution* 56 (1): 118-49.
- Fearon, James D., and David D. Laitin. 2000. "Violence and the Social Construction of Ethnic Identity." *International Organization* 54 (4): 845-77.
- Gerber, Elisabeth R., Adam Douglas Henry, and Mark Lubell. 2013. "Political Homophily and Collaboration in Regional Planning Networks." *American Journal of Political Science* 57 (3): 598-610.
- Griffiths, Ryan D. 2016. *Age of Secession: The International and Domestic Determinants of State Birth*. New York: Cambridge University Press.
- Gurr, Ted R. 1970. *Why Men Rebel*. Princeton, NJ: Princeton University Press.
- Gutiérrez Sanín, Francisco, and Elisabeth Jean Wood. 2014. "Ideology in Civil War: Instrumental Adoption and Beyond." *Journal of Peace Research* 51 (2): 213-26.
- Hafner-Burton, Emilie M., Miles Kahler, and Alexander H. Montgomery. 2009. "Network Analysis for International Relations." *International Organization* 63 (3): 559-92.
- Hoff, Peter D. 2015. *Dyadic data analysis with AMEN*. Accessed June 1, 2017. <https://arxiv.org/abs/1506.08237>.
- HooverGreen, Amelia. 2016. "The Commander's Dilemma: Creating and Controlling Armed Group Violence." *Journal of Peace Research* 53 (5): 619-32.
- International Crisis Group. 2012. *Tentative Jihad: Syria's Fundamentalist Opposition*. Accessed March 24, 2017. <https://d2071andvip0wj.cloudfront.net/131-tentative-jihad-syria-s-fundamentalist-opposition.pdf>.
- International Crisis Group. 2014. *Flight of Icarus? The PYD's Precarious Rise in Syria*. Accessed March 24, 2017. <https://www.crisisgroup.org/middle-east-north-africa/eastern-mediterranean/syria/flight-icarus-pyd-s-precarious-rise-syria>.

- Jones, Seth G. 2017. *Waging Insurgent Warfare: Lessons from the Vietcong to the Islamic State*. Oxford, UK: Oxford University Press.
- Kim, Nam Kyu. 2018. "Revolutionary Leaders and Mass Killing." *Journal of Conflict Resolution* 62 (2): 289-317.
- Krause, Peter. 2013/2014. "The Structure of Success: How the Internal Distribution of Power Drives Armed Group Behavior and National Movement Effectiveness." *International Security* 38 (3): 72-116.
- Krause, Peter. 2017. *Rebel Power: Why National Movements Compete, Fight, and Win*. Ithaca, NY: Cornell University Press.
- Lai, Brian, and Dan Reiter. 2000. "Democracy, Political Similarity, and International Alliances, 1816-1992." *Journal of Conflict Resolution* 44 (2): 203-27.
- Lichbach, Mark I. 1995. *The Rebel's Dilemma*. Ann Arbor: Michigan University Press.
- Lister, Charles R. 2015. *The Syrian Jihad: Al-Qaeda, the Islamic State, and the Evolution of an Insurgency*. New York: Oxford University Press.
- Maoz, Zeev. 2012. "Preferential Attachment, Homophily, and the Structure of International Networks, 1816-2003." *Conflict Management and Peace Science* 29 (3): 341-69.
- McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27:415-44.
- Metternich, Nils W., Cassy Dorff, Max Gallop, Simon Weschle, and Michael D. Ward. 2013. "Antigovernment Networks in Civil Conflicts: How Network Structures Affect Conflictual Behavior." *American Journal of Political Science* 57 (4): 892-911.
- Minhas, Shahryar, Peter D. Hoff, and Michael D. Ward. 2016. "A New Approach to Analyzing Coevolving Longitudinal Networks in International Relations." *Journal of Peace Research* 53 (3): 491-505.
- Minhas, Shahryar, Peter D. Hoff, and Michael D. Ward. 2019. "Inferential Approaches for Network Analysis: AMEN for Latent Factor Models I." *Political Analysis* 27 (2): 208-22. doi: 10.1017/pan.2018.50
- Morrow, James D. 1991. "Alliances and Asymmetry: An Alternative to the Capability Aggregation Model of Alliances." *American Journal of Political Science* 35 (4): 904-33.
- Oppenheim, Benjamin, Abbey Steele, Juan Vargas, and Michael Weintraub. 2015. "True Believers, Deserters, and Traitors: Who Leaves Insurgent Groups and Why." *Journal of Conflict Resolution* 59 (5): 794-823.
- Perry, Tom, and Suleiman Al-Khalidi. 2017. "Syrian Rebels Bitterly Divided Before New Peace Talks." Accessed May 14, 2018. <https://www.reuters.com/article/us-mideast-crisis-syria-rebels-analysis/syrian-rebels-bitterly-divided-before-new-peace-talks-idUSKBN1541NC>.
- Phillips, Christopher. 2016. *The Battle for Syria: International Rivalry in the New Middle East*. New Haven, CT: Yale University Press.
- Popovic, Milos. 2018. "Inter-rebel Alliances in the Shadow of Foreign Sponsors." *International Interactions* 44 (4): 749-76.
- Rudloff, Peter, and Michael G. Findley. 2016. "The Downstream Effects of Combatant Fragmentation on Civil War Recurrence." *Journal of Peace Research* 51 (1): 19-32.
- Salehyan, Idean. 2010. "The Delegation of War to Rebel Organizations." *Journal of Conflict Resolution* 54 (3): 493-515.

- Salehyan, Idean, David Siroky, and Reed M. Wood. 2014. "External Rebel Sponsorship and Civilian Abuse: A Principal-agent Analysis of Wartime Atrocities." *International Organization* 68 (3): 633-61.
- Seymour, Lee, Kristin M Bakke, and Kathleen Gallagher Cunningham. 2016. "E Pluribus Unum, Ex Uno Plures: Competition, Violence, and Fragmentation in Ethnopolitical Movements." *Journal of Peace Research* 53 (1): 3-18.
- Shesterinina, Anastasia. 2016. "Collective Threat Framing and Mobilization in Civil War." *American Political Science Review* 110 (3): 411-27.
- Staniland, Paul. 2014. *Networks of Rebellion: Explaining Insurgent Cohesion and Collapse*. Ithaca, NY: Cornell University Press.
- Staniland, Paul. 2015. "Militias, Ideology, and the State." *Journal of Conflict Resolution* 59 (5): 770-93.
- Tamm, Henning. 2016. "Rebel Leaders, Internal Rivals, and External Resources: How State Sponsors Affect Insurgent Cohesion." *International Studies Quarterly* 60 (4): 599-610.
- Ugariza, Juan E., and Matthew J. Craig. 2012. "The Relevance of Ideology to Contemporary Armed Conflicts: A Quantitative Analysis of Former Combatants in Colombia." *Journal of Conflict Resolution* 57 (3): 445-77.
- Walter, Barbara F. 2017. "The Extremist's Advantage in Civil Wars." *International Security* 42 (2): 7-39.
- Ward, Michael D., Katherine Stovel, and Audrey Sacks. 2011. "Network Analysis and Political Science." *Annual Review of Political Science* 14:245-64.
- Werner, Suzanne, and Douglas Lemke. 1997. "Opposites Do Not Attract: The Impact of Domestic Institutions, Power, and Prior Commitments on Alignment Choices." *International Studies Quarterly* 41 (3): 529-46.
- Wood, Reed M., and Jacob D. Kathman. 2015. "Competing for the Crown: Interrebel Competition and Civilian Targeting in Civil War." *Political Research Quarterly* 68 (1): 167-79.
- Zech, Steven T., and Michael Gabbay. 2016. "Social Network Analysis in the Study of Terrorism and Insurgency: From Organization to Politics." *International Studies Review* 18 (2): 214-43.

Appendix 3

Integrating Computational Modeling and Experiments: Toward a More Unified Theory of Social Influence

13

Integrating Computational Modeling and Experiments: Toward a More Unified Theory of Social Influence

Michael Gabbay

Applied Physics Laboratory, University of Washington, Seattle, WA 98105, USA

Introduction

Social influence is a central element of many behavioral areas, such as public opinion change, radicalization, and group decision-making – all of concern to public policy. It affects the process by which people form attitudes toward their governments, other population groups, or external actors. Group decision-making applications span political, military, economic, and legal domains. If computational social science is to aid in understanding, anticipating, and shaping such real-world contexts, then the development of accurate and broadly applicable models of social influence is essential. This chapter proposes that a deliberate and concerted integration of experimental investigation and computational modeling is needed to develop these models, an effort that will also advance the fundamental knowledge of social influence dynamics.

Within social psychology, social influence has largely been studied via the experimental testing of discrete theoretical hypotheses that express how a dependent variable responds to a change in an independent variable in qualitative language (e.g. increases, decreases, inverted U-shape). Quantitative reasoning is often employed in the rationale for theoretical propositions, and, to a much lesser extent, formal or computational models are also used to motivate them. While this process has been successful in revealing and extensively probing individual phenomena, it has been less effective at synthesizing and reconciling concurrent and competing processes. Such synthesis would better inform the development of computational models as to the relative strengths of different processes and their interaction. It is especially important if one seeks to apply a model to anticipate the behavior of a particular group of interest. For instance, the group polarization effect might predict that a group will pursue an extreme policy, whereas majority influence points toward a

moderate policy. Consequently, guidance as to how those two processes play out together is necessary in order to model the group's overall behavior. That guidance, of course, would need to be context dependent. As will be seen, the failure of group polarization theory to be integrated with broader social influence phenomena leads it to predict that every group of like-minded members will become more extreme, regardless of their initial opinion distribution. This is not so for the frame-induced polarization theory to be described below that does integrate group polarization with majority influence and consensus pressure: the theory can explain a systematic tendency for like-minded groups to become more extreme while being able to predict that individual groups will not.

The difficulty of synthesis in traditional social influence research has not, however, deterred a surge of modeling research across a range of disciplines. This activity has not been an unalloyed good for computational social science as much of this work has proceeded with little regard for empirical support. Sizeable and divergent streams of research have arisen around particular modeling approaches with murky domains of validity. This proliferation casts doubt upon the empirical relevance of the associated behavioral findings and complicates model selection and evaluation for applications.

The integrated approach advocated here calls for experiments designed with the explicit purpose of quantitatively testing computational models against data. It will help restore the balance between modeling and experimental validation. The development of computational models in conjunction with experiment will force researchers to reckon more intently on combining concurrent effects in order to make quantitative predictions. That is, a more unified theory will have effects be caused by multiple factors that earlier work associates with separate hypotheses. A greater orientation of experiments toward testing models rather than seeking new effects will encourage replication efforts and so place empirical findings on more solid ground. The increased focus on synthesis and the inevitable failures of previously successful models as they are tested in new regimes will spur theoretical innovation as well. Eventually, this integrated modeling-experiment approach will lead to convergence upon a set of social influence models that have substantial experimental support and so can be confidently extended to larger-scale systems or included within more complex simulations of particular application contexts.

This chapter first presents a brief survey of social influence and related group decision-making research, along with a discussion of how standard hypothesis testing and also quantitative modeling have been employed. The second section provides an overview of opinion network modeling. Next, the quantitative testing of computational models on experimental data is illustrated using recent work on group polarization (Gabbay et al. 2018; Gabbay forthcoming), which also shows how the modeling goal of synthesizing concurrent effects can lead to new and more unified theory. The fourth section then sketches the envisioned integration of modeling and experiment and discusses its potential benefits.

Social Influence Research

The term social influence is broadly applicable to both attitudinal and behavioral effects of human interactions. We focus upon research involving attitudes, opinions, and judgments here, mental constructs that often guide behavior. Classic social influence phenomena include conformity – the tendency and pressures toward consensus in groups, majority and minority influence – the ability of majorities and minorities to sway group opinions, and group polarization – the tendency of discussion among like-minded people to make positions more extreme. Two primary types of influence routes have often been invoked as explanations of such behaviors, normative and informational. Normative influence refers to the operation of group and broader social norms in setting expectations as to appropriate opinions and the value of consensus. Informational influence is the acceptance of information from others as evidence about the reality of the subject under consideration. Informational influence typically involves the alignment of one's private and publicly expressed judgments, whereas public agreement need not imply private acceptance under normative influence.

The body of social influence research above has been established through a process of hypothesis testing via laboratory experiments. Groundbreaking studies were conducted in the 1950s and 1960s (see Eagly and Chaiken 1993, Chapter 13). In a classic experiment by Asch, a substantial proportion of subjects suspended the clear evidence of their senses when faced with a majority of experimental confederates who stated that a clearly shorter line was longer, thereby demonstrating the power of conformity to induce public compliance. Inverting the direction of influence as the process under investigation, Moscovici and collaborators found in a color discrimination task that minorities who advocated consistent positions were more effective in swaying subjects than inconsistent minorities, leading to a theory that majorities primarily exert normative influence, while minority influence occurs mostly via the informational route. Group polarization, the tendency of discussion among like-minded individuals to lead to more extreme opinions, is another element in the social influence canon. It has both informational and normative influence explanations and will be discussed in detail below.

Research on group decision-making in contexts that allow for interpersonal persuasion also involves social influence. One strain of decision-making research considers the performance of groups in comparison with individuals (Kerr and Tindale 2004). For example, the wisdom-of-the-crowds hypothesis holds that simply aggregating individual judgments over many individuals yields greater accuracy than the judgments of individual experts under the assumption that the members of the pooled population make independent judgments whose uncorrelated errors cancel. Arguments have been made that social influence can either impair this performance by inducing correlated

errors or improve it when greater individual confidence tends to be associated with greater accuracy (Becker et al. 2017).

The vast majority of experiments on social influence and decision-making have been aimed at testing discrete qualitative hypotheses. A hypothesis is proposed concerning the direction of an effect on the dependent variable, increase or decrease, due to the variation of an independent variable, which is then tested statistically. If the amount of change in the dependent variable is in the theorized direction and improbably attributed to the null hypothesis of no effect, then the proposed hypothesis is said to be supported by the data.

Within the social influence and group decision-making domain, research affiliated with the literature on social decision schemes (SDS), which are essentially mathematical rules for combining group member initial preferences into a final decision, has most consistently pursued a model-based quantitative approach. Although initially concerned with juries and binary (innocent/guilty) decisions, the SDS program grew to include decisions concerning quantitative judgments such as monetary awards and budgets (Hinsz 1999). While it has been very successful with respect to its original jury concentration (Devine 2012), a shortcoming of the SDS program is that its focus on testing an array of aggregation rules has come at the expense of deeper theoretical and model development with respect to a specific opinion change process. This absence of a theoretical impetus inhibits generalization of the results to broader contexts – for example, when subgroup and social network structure is important or for general opinion dynamics in populations not associated with a focal decision point.

Opinion Network Modeling

Although the experimental study of social influence has been conducted by social psychologists and, to a much lesser degree, in other social sciences such as political science, sociology, and economics, the modeling of social influence dynamics has extended beyond social science to fields including physics, applied mathematics, computer science, and electrical engineering (Crandall et al. 2008; Castellano et al. 2009; Proskurnikov and Tempo 2017). While the primary goal of opinion network models is to predict final opinions from initial ones, the models typically describe a process that occurs over time. This section briefly discusses approaches to opinion network modeling and their empirical application.

Many models of opinion change have been developed in the fields noted above and beyond, involving a great diversity in methodological and substantive choices. One major methodological division involves whether outcomes are produced deterministically or stochastically. A fundamental substantive division involves the way opinions are mathematically represented. A binary

representation is clearly applicable to situations that ultimately involve a decision over two alternatives such as a political election. Alternatively, a continuous representation can account for gradations of opinion on an issue or for decisions involving either explicit numerical quantities such as budgets or that can be approximated as a spectrum of options ordered along some dimension (e.g. the extent of an escalatory military response). Binary (or discrete) opinion models tend to have stochastic interactions; continuous opinion models usually (but need not) employ deterministic interactions. The choice of opinion representation also constrains the basic process that governs how opinions change when nodes (a term for individuals within a network) interact. A binary opinion must either remain the same or flip when a node interacts with other nodes. In the voter model, a dyadic interaction is assumed whereby a node adopts the opinion held by a network neighbor selected at random, whereas the majority rule model proceeds by selecting subgroups at random with all member nodes adopting the majority opinion in the subgroup (Castellano et al. 2009). Continuous opinion models, on the other hand, allow for incremental shifts in opinion where the amount of change is a function of the distance between node opinions. The DeGroot and Friedkin–Johnsen models, as well as the consensus protocol (popular in the engineering literature on control), use a linear dependence in which the shift is proportional to the opinion difference (DeGroot 1974; Olfati-Saber et al. 2007; Friedkin and Johnsen 2011). Bounded confidence models assume a hard opinion difference threshold, within which nodes interact linearly but beyond which interaction produces no change (Lorenz 2007). The nonlinear model of Gabbay (2007) uses a soft threshold so that, rather than vanishing completely, the interaction decays smoothly with distance.

The vast majority of papers on opinion network models make no contact with empirical data. They start with a model, reasoned to be plausible (sometimes on the basis of social psychology research but sometimes on an appeal to common sense), and then generate simulation results, often in combination with mathematical analysis, on phenomena such as how the time to reach consensus scales with system size, the conditions conducive to the formation of camps of rival opinions, or the ability of extremists or influential individuals to shift opinions. Usually, the focus is on large systems taken to be representative of population-scale behavior. Consequently, such empirical connections as are reported are usually on the level of noting that model-generated curves exhibit qualitatively similar shapes to relationships observed in naturally occurring data from large population systems (Crandall et al. 2008; Düring et al. 2009; Török et al. 2013). However, some models have been shown to quantitatively reproduce empirical relationships such as the distribution of votes in proportional elections (Fortunato and Castellano 2007; Burghardt et al. 2016).

Application of opinion network models to laboratory experiments remains mostly confined to testing models developed within traditional fields of human

behavioral research rather than from the physical sciences and engineering. Friedkin and Johnsen (2011) conducted experiments in which they manipulated network topology for small groups and measured initial opinions, thereby enabling the quantitative comparison of experimental and model results. Although the communication topology was controlled, the network weights assessing interpersonal and self-influence in the model had to be calculated for each group separately on the basis of subjects' post-discussion ratings of interpersonal influence, thereby limiting predictive capability. However, their work remains the most extensive experimental investigation of an opinion network model. More recent work has employed agent-based modeling to qualitatively support and extend experimental results (Mäs and Flache 2013; Moussaïd et al. 2013, 2015).

Integrated Empirical and Computational Investigation of Group Polarization

This section provides an illustration of how experiment and opinion network modeling can be integrated as applied to group polarization and serves as a prelude to the description of the integrated approach in the next section. Recent research is described that demonstrates how a modeling-oriented approach can synthesize previously disjoint phenomena, generating a novel theoretical explanation of a classic social influence phenomenon, which furthermore predicts an effect unanticipated by existing theory (Gabbay et al. 2018). The basic theory is implemented in a simple aggregation model that integrates group polarization with the fundamental social influence processes of majority influence and conformity. Further, a model of opinion network dynamics shows how this basic process can arise from a lower-level attitude change framework that considers how persuasive messages shift both opinion and its associated uncertainty (Gabbay forthcoming). Both models not only qualitatively agree with the results of an online discussion experiment but, in accord with the proposed integration of modeling and experiment, are in quantitative agreement with the data as well.

Group Polarization Theory

The question of how groups shift toward more extreme positions has been a focus of both traditional social influence research and opinion network modeling although the explanatory mechanisms favored by each are disconnected. Group polarization is said to occur when, in a group composed of individuals already on the same side of an issue, the post-discussion mean opinion shifts further in support of that side as compared with the pre-discussion mean (Myers 1982; Brown 1986; Isenberg 1986; Sunstein 2002). Note that, contrary

to common parlance, *polarization* here refers to movement toward one pole rather than divergence toward opposite poles. The seminal experiments in the 1960s focused on *choice dilemmas* in which subjects were presented with hypothetical scenarios involving the choice between a risky but higher payoff option over a safer, lower payoff one (Brown 1986). Subjects were asked to choose the minimum odds of success they would accept in order to pursue the riskier option. For most choice dilemma items, discussion led groups to choose lower odds of success as measured by the difference in the group pre- and post-discussion means. The effect, therefore, was originally coined the *risky shift*. However, some choice dilemma items tended to produce shifts toward greater caution, while others produced no shift in either direction. Cautious items were marked by a very large stake such as someone's health or marriage, whereas risky items tended to offer a large potential gain for a small stake. Experiments on group betting involving real rather than hypothetical stakes also have shown a mix of risky and cautious shifts (Isenberg 1986).

Beyond the risk context, discussion among similarly inclined individuals was found to cause more extreme social and political attitudes (Myers and Bishop 1970; Schkade et al. 2010; Keating et al. 2016). Manipulation of the evidence presented to mock juries exhibited discussion-induced shifts to lower presumed guilt and softer sentences in cases where the evidence was weaker and higher presumed guilt and harsher sentences for stronger evidence (Myers 1982). Similarly, jury damage awards exhibit polarization (Schkade et al. 2000). In general, the contexts in which group polarization occurs are on the judgmental side of the intellective–judgmental spectrum in which purely intellective tasks, such as math problems, have demonstrably correct solutions, whereas purely judgmental tasks are matters of personal taste or aesthetics (Laughlin and Ellis 1986). Most real-world decision contexts such as forecasting and policy-making are characterized by both intellective and judgmental aspects; they may draw on a body of knowledge (e.g. expertise on a country's political system), yet judgments must be made as to significant uncertainties (e.g. the intentions of political leaders).

Corresponding to the two main pathways of social influence, informational and normative processes underlie the two main explanations of group polarization within social psychology. In the informational route, known as persuasive arguments theory, group members expose each other to new information in favor of that side. In the normative route, known as social comparison theory, a group norm associated with the broader culture or that particular group's identity defines a preferred direction on an issue so that opinions shift in the direction of the norm; a norm toward risk-taking, for instance, would lead groups to make riskier choices as a result of discussion. While the informational and normative mechanisms for group polarization have received robust experimental support, they have never been integrated with

strong social influence phenomena such as consensus pressure and majority influence. Relatedly, neither explanation has been developed with respect to a clear formal model at the individual group level. Although the informational and normative processes occur at the group level, these theories were operationalized mainly with respect to a population of groups with random initial opinion distributions, over which majority influence could be assumed to cancel out. As a result, group polarization theory is effectively silent as to whether a particular group with a specific initial distribution of opinions will become more extreme. Alternatively, one could make a strong reading of either persuasive arguments theory or social comparison theory that neglects other processes in which they always predict polarization for homogeneous groups (for sufficiently judgmental issues). Either alternative – silence or a uniform prediction of polarization – limits the ability of existing group polarization theory to address real-world contexts such as whether, in the face of a foreign policy crisis, discussion among a country's leadership will induce a shift toward a more extreme course of action.

Opinion network models do not suffer from an inability to go from initial to final opinions since that is their fundamental purpose. The dominant approach to modeling extremism within this literature has been to attribute higher network weights to nodes with more extreme initial opinions (Deffuant et al. 2002; Friedkin 2015). This *extremist-tilting* approach is necessitated by the fact that in most continuous opinion models the mean opinion in networks with symmetric weights (i.e. the strength of influence is the same from node i to j as from node j to i) remains constant at its initial value, therefore preventing the shift in mean exhibited in group polarization. Consequently, extremists must be assigned greater influence over moderates than vice versa in order to shift the mean. Psychologically, this move is attributed to extremists' greater confidence, commitment, or stubbornness. Extremist tilting is not widely accepted within the literature dedicated to group polarization, however, and has received only mixed experimental support (Zaleska 1982).

Frame-Induced Polarization Theory

This section discusses the frame-induced theory of group polarization introduced by the author and Zane Kelly, Justin Reedy, and John Gastil (Gabbay et al. 2018). This theoretical mechanism is complementary to and can operate simultaneously with the mechanisms of standard polarization theory (shorthand for both persuasive arguments theory and social comparison theory). However, the frame-induced mechanism provides an explanation of group polarization that, unlike standard polarization theory, is integrated with consensus pressure and majority influence, thereby enabling prediction given the group initial opinion

distribution. The theory is developed specifically with respect to a quantitative policy under debate, although it should prove applicable to opinions more generally. Examples of quantitative policies include budgets, investment amounts, interest rates, jury damage awards, or military operation sizes. In its emphasis on how the policy is discussed, the theory takes into account the basis of policy positions and not just the policy value alone. For the particular context discussed here, this basis is grounded in the theory of decision-making under risk and uncertainty (Pleskac et al. 2015) and so constitutes a further theoretical element that is integrated within the frame-induced theory.

Crucial to the frame-induced theory is the recognition of a distinction between the policy under debate and the *rhetorical frame* by which it is discussed. In general, one would expect the rhetorical frame to be a substantive aspect of the policy for which there is substantial disagreement among group members. The frame could represent a key uncertainty or differences in how group members value the outcomes associated with the policy. A given issue may admit multiple frames if there are different dimensions of comparable disagreement. Yet, a single dominant rhetorical frame may emerge due to group-specific dynamics such as deliberate efforts to focus a debate as occurs in political framing (Chong and Druckman 2007). Persuasion, and hence agreement, is driven by proximity along the rhetorical frame not the policy itself. The shape of the *rhetorical function* that maps policy positions into positions along the rhetorical frame (*rhetorical positions*) plays an essential role in frame-induced polarization theory.

Focusing on when uncertainty is the source of the rhetorical frame, uncertainty can generate disagreement if group members have different estimates of the probability of either an unknown variable or an impending outcome. A simple but important example involves a policy that depends on the outcome of a binary gamble so that the likelier one estimates the outcome to be: the more stake one is willing to risk on its occurrence. For instance, one would prefer to invest more in a defense technology company (the policy), the greater one's subjective probability that the pro-defense spending candidate in an election is likely to win (the rhetorical frame). The use of the subjective probability of a binary outcome as the rhetorical frame is also relevant to the experiment described below.

Two important behaviors that impact group polarization arise from the distinction drawn between policy and the rhetorical frame: (i) *distribution reshaping*, which preferentially facilitates the formation of extreme majorities and so generates group polarization, and (ii) *heuristic frame substitution*, which can enhance polarization on one side of the issue and suppress it on the other. Distribution reshaping arises when a nonlinear rhetorical function causes the relative spacing between group member rhetorical positions to be

different than between their policy positions. Consequently, the distribution of group member rhetorical positions will be reshaped with respect to the distribution of policy positions. Such reshaping may reduce the rhetorical distance within some subgroups relative to others as compared with their distances directly along the policy itself, thereby affecting the composition of the majority that emerges during deliberations. Specifically, for a concave (downward curvature) rhetorical function, rhetorical position increases more slowly as the policy becomes more extreme. This causes the rhetorical distance between more extreme members to be compressed relative to the distance between more moderate ones. This compression favors the emergence of a majority at the extreme, which then drives consensus to a policy that is more extreme than the mean of the initial policy distribution. For the case where the policy, such as a wager or investment, arises from the subjective probability of an outcome in a binary gamble, Gabbay et al. (2018) show that a concave rhetorical function is expected using the theory of decision-making under risk and uncertainty (Pleskac et al. 2015).

As an illustration of distribution reshaping, consider a group of three military planners in wartime tasked with deciding whether to increase or decrease the size of the force allocated to defend a certain territory. The policy is then the change in the number of troops, positive or negative, from the current level. Take a planner's preferred force level to be a function of their estimate of the probability of an enemy offensive against this territory and their assessment of its worth relative to other territories. If there is little disagreement as to worth yet there is fundamental uncertainty as to enemy intentions, then the subjective probability of an enemy offensive is expected to be the dominant source of disagreement and, hence, the rhetorical frame. The rhetorical function is then the transformation that maps a given change in force level to the corresponding subjective probability of an enemy offensive. For instance, assume that the three planners are all inclined to boost the force level and that their respective preferences for the increase in troops are (500, 1500, 2500), which are mapped by the rhetorical function to subjective probability estimates of (0.55, 0.65, 0.70). These values indicate that the rhetorical function is nonlinear: a policy difference of 1000 between the first two members corresponds to a change in probability of 0.10, while a 1000 difference between the second and third members yields a probability change of only 0.05. More specifically, it is concave in that the subjective probability goes up at a slower rate as the increase in force level becomes more extreme. While the policy distribution consists of one member at the mean of 1500 with the other two an equal distance below and above it, the rhetorical position distribution has one member below its mean of 0.633 and two above; a symmetric policy distribution has been reshaped into an asymmetric rhetorical one. As will be described presently, the theorized opinion change process assumes that the two more extreme members are likely to form a

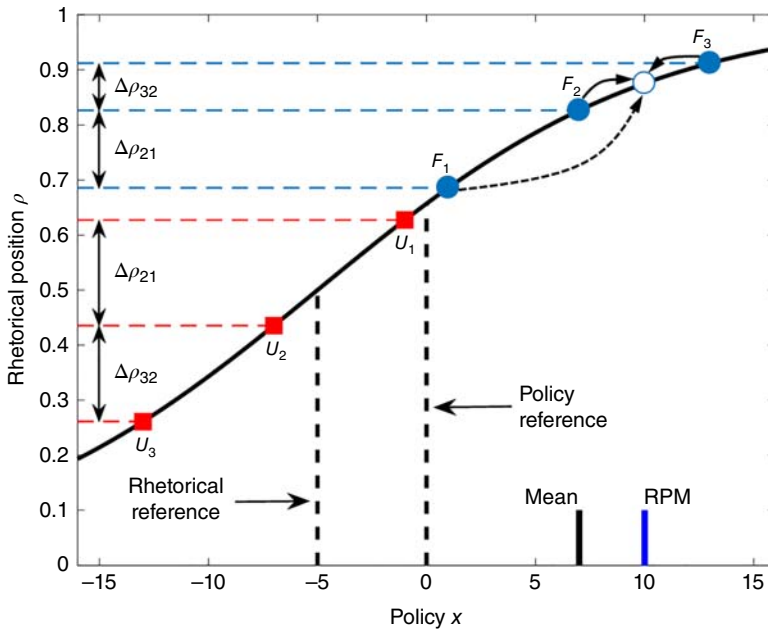


Figure 13.1 Illustration of distribution reshaping, RPM process, and reference point shifting due to heuristic frame substitution. Rhetorical function relating rhetorical frame position to policy is solid curve. Solid curved arrows indicate formation of the RPM pair by F_2 and F_3 ; dashed curved arrow indicates conformity of F_1 to form final consensus. Short dark gray line at the bottom is the mean of the initial F group policies; short light gray line is RPM process prediction for the F group consensus policy.

majority, converging at their midpoint of 2000, which yields the ultimate group consensus.

To yield a systematic tendency for homogeneous groups to shift toward the extreme, distribution reshaping must be linked to an opinion change process. The rhetorically proximate majority (RPM) process specifically treats consensus formation. Most group polarization experiments have required consensus, and it was the most common outcome in our experiment. Figure 13.1 illustrates how distribution reshaping combines with the RPM process to generate group polarization. Two separate three-person groups, the F and U groups (the reason for these designations will be revealed below), are depicted. The policy reference at $x = 0$ is the boundary between the opposing policy sides. If a group is entirely on one side of the policy reference, then it is said to be homogeneous. The F and U groups are seen to be homogeneous on the positive (pro) and negative (con) policy sides, respectively. Both groups have symmetric policy distributions given by $(x_1, x_2, x_3) = (\pm 1, \pm 7, \pm 13)$ where the $+$ ($-$) sign corresponds to the F (U) group.

The F group is used to describe distribution reshaping and the RPM process. The rhetorical function $\rho(x)$ maps the policy x to the rhetorical position ρ . The rhetorical position scale goes from 0 to 1 and so can correspond to the choice of subjective probability as rhetorical frame. Although the policy of the centrist F_2 is exactly midway between those of the moderate F_1 and the extremist F_3 , the F group members are arrayed along the shoulder of the rhetorical function so that, with respect to the rhetorical frame, F_2 is closer to F_3 than to F_1 ; the rhetorical distance $\Delta\rho_{32}$ is much less than $\Delta\rho_{21}$. As agreement is driven by proximity of rhetorical positions, it is therefore more likely that F_2 will join with F_3 to form the RPM pair than with F_1 . There remains the question of where the RPM pair will converge. Assuming F_2 and F_3 have equal influence on each other, then they should converge at either the midpoint between their rhetorical positions, which is then transformed back to the corresponding policy position, or midway between their policies. Either choice will lead to a policy more extreme than the mean x_2 , but the latter process is more direct than the former. In addition, the policy positions are explicitly numerical, whereas the rhetorical ones need only be expressed qualitatively. It is therefore more plausible that the RPM pair forms midway between their policies (at $x_{23} = (x_2 + x_3)/2 = 10$). The convergence of the RPM pair on this point is indicated by the solid arrows leading to the open circle. As indicated by the dashed arrow, majority influence then causes F_1 , now in the minority, to conform to the F_2, F_3 position. The result is a consensus policy that is more extreme (i.e. further from the policy reference) than the initial policy mean – in accord with definition of group polarization.

Although distribution reshaping due to a concave rhetorical function in combination with the RPM process can explain group polarization, it predicts only a systematic tendency for groups to shift toward the extreme. It is clear from Figure 13.1 that if F_2 's initial policy were moved sufficiently close to F_1 , then those two would form the RPM pair, which would yield a consensus policy below the mean. This ability to predict that individual groups can depolarize against an overall polarization tendency stands in contrast to standard theory's strong prediction of polarization for every homogeneous group.

Note that the references corresponding to the policy and rhetorical frame are offset along the horizontal axis in Figure 13.1. This misalignment between the policy and rhetorical references can arise from heuristic frame substitution in which a simpler heuristic rhetorical frame is discussed in place of a more complex frame that directly corresponds to the policy. In the binary gamble context, such substitution can occur when there are two distinct gambles that depend on the same random variable but with different thresholds: the policy gamble that directly determines whether one's policy choice is successful and a heuristic gamble that is more intuitively accessible. In a stock investing scenario, for example, the policy gamble could be whether or not the return of the stock over a given period of time would exceed that of a fixed return asset such as a bond.

If the stock's return is greater than the bond's return, then the investment is successful. The proper rhetorical frame would then be the subjective probability of that outcome. If it is greater than 0.5, then one should invest in the stock and not the bond. However, discussions about the investment might focus on the more intuitive heuristic gamble of whether the stock's price will rise or fall. Both of these gambles depend on the same random variable, the stock's return, but with different thresholds – the fixed return for the policy gamble, zero for the heuristic gamble – and so they have distinct subjective probabilities. If a probability of 0.5 is taken as the neutral reference for both subjective probabilities, then they will be related to different policy reference points by the rhetorical function.

The U group in Figure 13.1 illustrates the effect of reference point shifting due to heuristic frame substitution. The reference point of the rhetorical function is shifted left from that of the policy itself so that the U group members, who are all on the same (negative) side of the policy axis, straddle the rhetorical reference point (U_1 is to the right, while U_2 and U_3 are to the left). Because they are arrayed along the roughly linear part of the curve, the U group members are subject to weak distribution reshaping. One would expect therefore that the U group would be less prone to polarize than the F group. Strict application of the RPM process, however, would, for the case illustrated, lead to the formation of a U_2 and U_3 majority as the rhetorical distance $\Delta\rho_{32}$ is slightly less than $\Delta\rho_{21}$, which would then yield a substantial shift toward the negative extreme. But a small shift in rhetorical positions due to noise or uncertainty could readily flip which distance is smaller, causing the (U_1, U_2) majority pair to form, which would lead to depolarization instead. Whether the U group will polarize is consequently much harder to predict than for the F group. Considering a population of similar U groups, about equal numbers will become more moderate as become more extreme and so there will be no systematic polarization. In contrast, the offset of the rhetorical reference places the F group further along the shoulder and so enhances systematic polarization on the positive policy side.

Accept-Shift-Constrict Model of Opinion Dynamics

Two mathematical models have been presented in connection with frame-induced polarization theory. The RPM process described above is formulated as the RPM model, which determines a consensus policy by a weighted average of the policies of the majority of group members whose rhetorical positions span the least range (Gabbay et al. 2018). Network structure is accommodated by weighting policies by relative node degrees. Rather than static aggregation, the second model describes the opinion change process over time as a result of dyadic-level interactions. This accept-shift-constrict (ASC) model makes two innovations beyond existing continuous opinion network models (Gabbay forthcoming). First, it makes a distinction between

policy (or opinion, more generally) and rhetoric in accord with the theory above. Second, it incorporates a novel uncertainty reduction mechanism that does not require that node uncertainties be visible to others.

The ASC model assumes an underlying dyadic process in which one node sends a message to a receiver node in an effort to persuade the latter. The message can impact both the receiver node's policy and its uncertainty interval around that policy. Conceptually, the model proceeds in distinct *accept*, *shift*, and *constrict* phases (although all occur simultaneously in the mathematical formulation). The accept and shift phases occur in the equation that governs the rate of change of the node's policy. In the accept phase, the ASC model assumes that the probability that the receiver node will accept the message as persuasive decreases as a Gaussian function of the *rhetorical* distance between the sender and receiver nodes. The uncertainty of the receiver's position is taken to be the standard deviation parameter in the Gaussian. If a message is accepted, then, in the shift phase, the receiver shifts its policy in the direction of the sender's by an amount proportional to their *policy* difference. The constrict phase is governed by a second equation for the rate of change of a node's uncertainty, modeling a process in which interaction with others with close positions reduces uncertainty. If the sender's rhetorical position is within the uncertainty interval of the receiver, then the receiver decreases its uncertainty but not below a certain minimum value. Accordingly, unlike other models that involve uncertainty dynamics (Deffuant et al. 2002), it is the difference in (rhetorical) positions among dyad members rather than their difference in uncertainties that drives uncertainty change. The network weights in the ASC model represent the influence of one node upon another due to factors such as communication rate and expertise; they need not be symmetric. The ASC model is implemented in terms of coupled nonlinear ordinary differential equations, with two equations for each group member, one for the policy and one for the uncertainty.

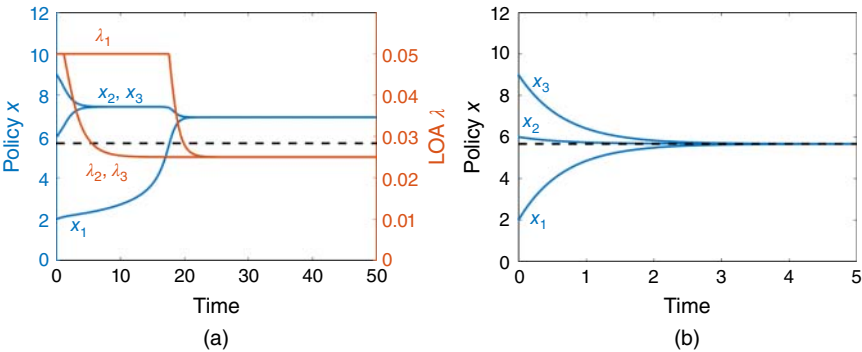


Figure 13.2 Evolution of policy positions and uncertainties for completely connected triad. (a) ASC model. (b) Consensus protocol. Dashed lines indicate initial mean policy.

A sample simulation of the ASC model as applied to a triad is shown in Figure 13.2a. All the nodes are connected and the network weights between nodes are symmetric and all equal. The initial policies are set such that x_2 is closer to x_3 than x_1 . We observe that x_2 and x_3 form a majority and their uncertainties (λ_2, λ_3) quickly reach their minimum values, while that of the minority member (λ_1) stays at its initial value. Consequently, x_1 is more open to accepting messages from the majority pair than vice versa, so that x_1 essentially comes up to the majority position, resulting in a consensus policy that is shifted upward from the mean. This ability for a majority to emerge and persist in the face of minority influence is a crucial part of the frame-induced mechanism of group polarization. This dynamic is not present in linear opinion network models such as the DeGroot, Friedkin–Johnsen, and consensus protocol models. As shown in Figure 13.2b for the consensus protocol (also known as the Abelson model, a continuous time equivalent of the DeGroot model), no interim majority emerges as all nodes converge simultaneously on the initial mean.

Experiment and Results

In the experiment reported in Gabbay et al. (2018), triads of knowledgeable fans of the National Football League (NFL) wagered on the outcomes of upcoming NFL games. In accordance with standard NFL betting practice, the wager did not concern simply which team would win the game but rather the margin of victory. Professional oddsmakers set an expected margin of victory, known as the point spread, by which the favorite team (the team deemed likely to win the game) is expected to win over the underdog. A bet on the favorite is successful if the favorite wins by more than the spread; otherwise, a bet on the underdog is successful (neglecting actual practice of returning bets when the favorite's victory margin equals the spread). The spread is set so as to estimate the median margin of victory if the game were repeated many times and, empirically, the chances of favorites and underdogs with respect to the spread are effectively equal. As a consequence, the payoff is the same regardless of which team one bets on.

Several days before the selected game, subjects drawn from an online labor pool were asked in a survey to choose which team they expected to win against the spread and how much they would wager from \$0 to \$7 on their choice. Triads were then assembled into online discussion groups according to three manipulated variables: (i) policy side conditions of *favorite* or *underdog* corresponding to the team that all group members had chosen (there were no groups of mixed team choice), (ii) disagreement level conditions of *high* (\$7) and *low* (\$3, 4, 5) corresponding to the difference between the highest and lowest wagers in the group, and (iii) network structure conditions of *complete* in which all group members could send messages to each other and *chain* in which the intermediate wager person was the middle node and the low and high wager

individuals were the ends. Groups discussed the game for 20–30 minutes after which members selected their final individual wagers. The winnings from successful wagers were donated to a specified charity.

Analyzing only groups that made consensus wagers (the vast majority of outcomes), the polarization metric is the difference between the mean post- and pre-discussion wagers of each group. If the post-discussion mean is higher, the group is said to have displayed a risky shift. Statistically significant results were found for all three manipulated variables: (i) favorite groups displayed a large risky shift, whereas underdog groups showed a small shift not statistically distinguishable from zero; within the favorite groups, a greater risky shift was observed for (ii) high disagreement groups than low and (iii) complete networks than chains.

These results are not readily explained by either standard polarization theory or existing opinion network models. The differential polarization behavior in particular stands at odds with the informational and normative explanations, which predict that a risky shift should occur for both the favorite and underdog conditions. Since both favorite and underdog groups were homogeneous with respect to policy side, group members have more novel arguments in support of their team choice, and so persuasive arguments theory predicts polarization for both favorite and underdog groups. Similarly, given the low stakes of the task, social comparison theory predicts that a norm favoring risk taking should be present in both groups and so both should exhibit a risky shift. This differential polarization result is also counter to the extremist-tilting explanation of opinion network modeling because, presumably, a high bet on the underdog is equally as extreme as the same amount bet on the favorite and so the level of extremist tilting is the same for both sides. With respect to the results for disagreement and network structure, their effects have been under-theorized and under-explored in the literature. The only previous experimental investigation of network structure in group polarization found no effect of topology (Friedkin 1999).

The experimental results, however, are in qualitative agreement with frame-induced theory. The differential polarization by policy side arises from heuristic frame substitution in which the question directly related to the wager policy – who will beat the spread? – is replaced by the heuristic one – who will win the game? The subjective probability of the favorite winning the spread is the proper rhetorical frame but is replaced by the subjective probability of the favorite winning the game. While professional gamblers may be able to think directly in terms of the spread victor probability, the game victor probability is a much more natural one for most knowledgeable fans to consider and so constitutes the rhetorical frame operative in the discussion. This substitution also entails a shift of reference point since both gambles depend on the same random variable, the margin of victory, but with different thresholds for their resolution. The reference margin of victory for the spread victor gamble is the

point spread, whereas the game winner gamble has a reference of zero points. These different references for the margin of victory yield different policy and rhetorical reference wagers. The rhetorical reference is obtained by considering the wager for a subjective probability of the favorite winning the game equal to 0.5. Believing that the game is a toss-up implies that one estimates the margin of victory to be zero and so one should bet on the underdog if the oddsmakers have set a nonzero spread. Therefore, the rhetorical reference equates to some wager on the underdog. If positive and negative wagers are used to represent favorite and underdog bets, respectively, then the rhetorical reference corresponds to a negative wager. Accordingly, the F and U groups in Figure 13.1 can now be seen as analogous to the favorite and underdog groups in the experiment.

That favorite groups with high disagreement are expected to show a greater risky shift is a consequence of the RPM process described in connection with Figure 13.1. Expanding the difference between x_1 and x_3 while keeping x_2 fixed implies that the F_2 and F_3 will form the RPM pair at a more extreme policy since x_3 is more extreme. The greater polarization for complete networks vs. chains is due to the greater relative communication rate of the center node in the chain along with its intermediate wager. Rather than forming at their policy midpoint as in the complete network, the greater influence of the chain center node causes the RPM pair to form at a policy that is closer to x_2 and so implies a lesser shift to the extreme.

As a visual comparison between the data and models, Figure 13.3 displays the observed and simulated pre- to post-discussion shifts in the group mean wager as a function of the wager difference (averaged over all groups at each difference) where favorite and underdog groups are shown, respectively, on the positive and negative sides of the horizontal axis. Groups are simulated using their actual wagers and spreads. The weights in the complete and chain networks are set by a priori considerations of the topological effects upon communication rates. In the complete network, all weights are equal, whereas the middle node in the chain has twice the weight of the end nodes (these expectations are in approximate agreement with the measured communication rates). The free parameters in the models, the level of risk aversion plus the initial and minimum uncertainties (for ASC), were chosen so as to minimize the total χ^2 error over both networks between the observed and simulated data. The data displays the observed greater polarization for favorite groups, high disagreement level, and complete topology. The RPM and ASC simulations also display these behaviors demonstrating qualitative agreement between the experiment and simulations. That the simulation results mostly pass through the error bars further suggests quantitative agreement whose testing we now discuss.

In general, when statistically testing the fit of a model, one assesses whether its predictions are consistent with the data in the sense that it is reasonably probable that the model could have produced the data given the presence of

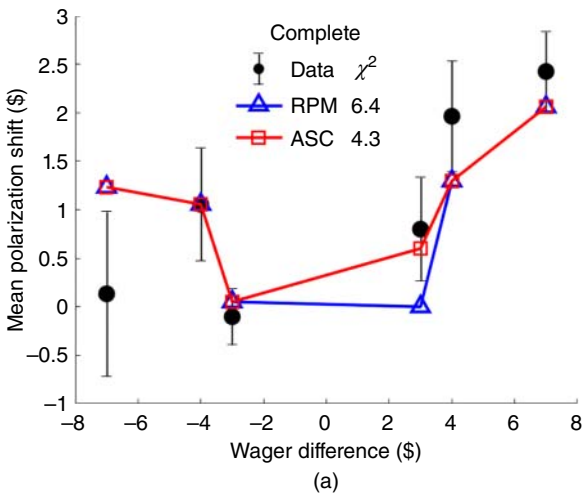
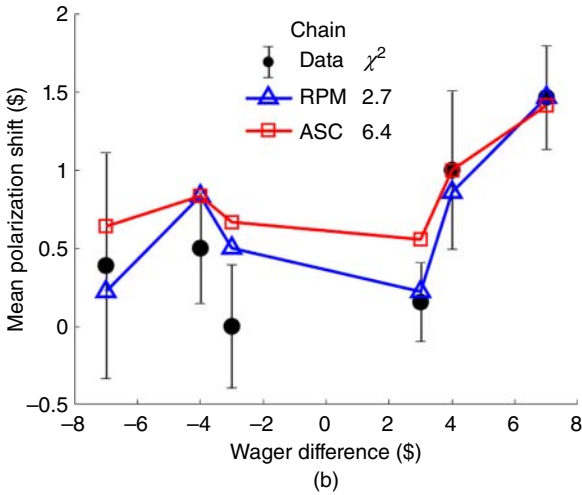


Figure 13.3 Comparison of experimental data and simulations of ASC and RPM models. (a) Complete network. (b) Chain. Positive and negative sides of the x -axis correspond to favorite and underdog groups, respectively. Error bars are standard errors. Simulation results are rounded upward to nearest dollar.



unmodeled noise. The null hypothesis is that the model is correct, and so support for the model is found if the null hypothesis cannot be rejected. Note that this criterion is opposite to that used in the testing of qualitative effects where one seeks to reject the null hypothesis in order to claim support for the theorized relationship. If the model passes the test, one can only claim that it is consistent with the data, not that it is the true model, whereas failure to pass the test indicates that the model can be rejected as false. Free parameters – those that cannot be determined without using the dependent variable – are fit so that the error statistic is minimized. Using more free parameters, however, has the effect of decreasing the maximum allowable error beneath which one can claim support for the model. This makes it harder for models with more

parameters to pass the test. However, if two models – a parameter-lean one and a parameter-rich one – both pass the test, one cannot claim more support for one than the other on the basis of the test itself. While the parameter-lean model might be preferred by virtue of its parsimony, the parameter-rich model may be preferred if it has more general scope beyond the experiment under study or if there are more grounds for its causal process in the relevant literature.

The RPM and ASC models were evaluated using a χ^2 goodness-of-fit test. A χ^2 goodness-of-fit test uses the sum of the squared errors between the observed and predicted data points (normalized by the variance at each point) as its test statistic. A threshold of $Q \leq 0.2$ was chosen for rejecting the null hypothesis that the model is correct. This threshold is conservative with respect to the standard significance threshold (p -value) of 0.05 used in testing qualitative hypotheses in that a higher Q -value makes it more difficult for the model to pass the test. The RPM model (one free parameter) has $Q = 0.61$, and the ASC model (three free parameters) yields $Q = 0.3$. Accordingly, both models were found to be consistent with the data. On the other hand, several alternative models, such as the median and a proximate majority model based directly on the policy (not the rhetorical frame), did not pass. Consequently, this test provides a statistical basis for rejecting these models as explanations of the experimental data.

While the agreement between the experiment and the models is encouraging, more experimentation is needed to judge the validity of the theory. Of greater relevance for present purposes is that this work illustrates some of the themes of the integrated approach. First, it demonstrates that quantitative agreement between computational models and experiment is possible. Plots such as Figure 13.3 comparing data points with error bars against model predictions as a function of an independent variable are common in the physical sciences, but not so in social influence research or social science more generally. The work also shows that the modeling goal of predicting outcomes for groups with specific initial conditions drives a heightened concern for the synthesis of concurrent effects, as done here for group polarization and majority influence. This concern for synthesis can lead to a new theory, which, in turn, can predict fundamentally novel phenomena such as differential polarization due to frame substitution. In addition, the synthesis driven by modeling can yield clear predictions for the effects of variables left ambiguous in qualitative theory, such as disagreement level and network structure.

Integrated Approach

The proliferation of opinion network modeling research that is largely uncoupled from empirical support may be impeding rather than advancing

computational social science. Particularly in fields outside of social science such as physics, computer science, and engineering, an initial model can be subjected to ever more sophisticated analysis or variation, thereby achieving prominence incommensurate with its level of empirical support. This proliferation makes it hard for application-oriented researchers to choose among various models as model popularity need not indicate empirical validity. Likewise, it is difficult for prospective operational consumers of simulations that incorporate opinion network models to evaluate their validity.

For its part, social psychology is not wanting of experimental investigation of social influence. Commonly, a new behavior is discovered experimentally and further experiments are aimed at generalizing and elaborating upon the origins and circumstances of its occurrence. Experiments can extend the behavior to new contexts and winnow down competing theoretical explanations, but much less effort is dedicated to synthesizing it with other social influence phenomena. This lack of synthesis hampers application to real-world contexts. Many experiments have been conducted on group polarization since the 1960s, but the results have not been firmly integrated with majority influence, consensus pressure, or attitude change research bearing on disagreement. As applied to a natural group decision-making context for an issue with a substantial judgmental element, the normative influence theory, for example, would advise identifying whether a culturally salient norm is present that would push groups further to the extreme. But then, given such a norm, it would always predict polarization. Another example is how the hidden profiles paradigm (Schulz-Hardt and Mojzisch 2012) has not been reconciled with the persuasive arguments theory of group polarization: the former emphasizes how group discussion centers on pieces of information held in common to the detriment of the sharing of unique information held by individuals, while the latter hinges upon group members exchanging their unique information.

An approach that centers on integrating experiment with the quantitative testing of computational models of social influence will place models on more solid empirical footing. Of particular importance to this approach are complex systems models (e.g. the ASC model), in which the variables of interest, such as group member opinions, evolve from initial conditions as a result of endogenous feedback with each other and perhaps exogenous signals (Gabbay 2014). Unlike standard statistical models, complex systems models do not directly posit some generic functional form, often linear or quadratic, that expresses an overall relationship between independent and dependent variables. Rather, the overall functions relating input and output are determined by the unfolding of the processes in the model and so need not result in a function that is expressed in a simple analytic form. However, the predicted relationship is more specific than simply saying, for instance, that there is an interaction effect between variables as in regressions. That complex systems models must deal with specific

group initial conditions, rather than an overall population of groups, provides a greater motivation to synthesize different effects than do statistical approaches.

Group polarization research provides an example of how seeking to model specific initial conditions can drive synthesis. One can statistically analyze the overall group polarization in a population of homogeneously inclined groups with random initial preference distributions without worrying about majority influence. Since groups with a majority preference above the mean would be roughly balanced by those whose majority was below the mean, (preference-based) majority influence could not be the cause of any observed net shift toward the extreme. However, majority influence strongly affects whether a particular group will polarize or not as it can operate counter to polarization when a group has a majority below the mean. This implies that majority influence cannot be ignored when one seeks to predict whether individual groups will polarize. Moreover, the focus on making predictions for specific within-group initial preference distributions, rather than over a population of groups with random distributions, helped spur the frame-induced theory's reconceptualization of majority influence as operating over the rhetorical frame, thereby becoming integral to the group polarization process.

Implementation of an integrated approach will entail changes for both the experimental and modeling sides. For the former, the major change is that experiments will be designed from the outset with the objective of quantitatively testing computational models, not just qualitative hypotheses. For hypothesis testing, one often coarse grains the values of an independent variable in order to construct experimental cells corresponding to, say, a binarization of that variable, as we did with disagreement level above. Quantitative testing of computational models, however, typically requires greater resolution, and experiments should be designed so that a sufficient number of variable values are present to conduct a goodness-of-fit test. As complex systems models can utilize specific experimental initial conditions for individuals or groups instead of treating them as random error, greater attention should be given to the distribution of initial conditions in the design than is needed when the goal is simply to populate binarized condition cells. Beyond a focus on initial and final states, the testing of complex systems models would also benefit from experimental measurements over time, when such measurement is feasible and does not unduly interfere with the process under study.

For the modeling side, a greater focus on developing models capable of being estimated from experimental data is needed. Making quantitative contact with experimental results requires more disciplined consideration of parameters than when more simply endeavoring to show a qualitative correspondence with the data. The temptation to develop rich models must be tempered against the need to estimate parameters from the data itself if they cannot be independently determined. Discretion should be exercised when considering the addition of parameters beyond those directly related to the variable of

primary concern. The essential parameters in continuous opinion network models will involve distance in the opinion space.

Another important element is that models be capable of prediction. This requires that, once parameters are estimated on an in-sample population, models are then capable of predicting cases held out of the sample or from a new experiment. As noted above, the network weights in the Friedkin–Johnsen model have primarily been calculated using post-discussion ratings of influence by the group members themselves. Such a procedure precludes prediction. However, it would be possible to test the extremist-tilting explanation of group polarization in conjunction with the Friedkin–Johnsen model if one were to fit the function relating persuasion resistance to opinion extremity.

The focus on a relatively tight number of parameters that can be determined a priori or estimated from the data will make models more robust and applicable across experiments falling into the same broad context. A more ambitious goal of an experimentally oriented modeling program would be to bridge different experimental contexts, such as problem solving, forecasting, policy-making, and ideological attitudes. As an example, the distinction between intellective and judgmental tasks is an important one in group decision-making. At opposite ends of the intellective–judgmental spectrum, purely intellective tasks like math puzzles have solutions that are demonstrably correct, whereas purely judgmental tasks are matters of personal taste. Forecasting problems, for instance, lie in between, having not only intellective elements that are demonstrably right or wrong, such as the record of a football team or what party has the most registered voters in a given district, but also judgmental ones involving the factors likely to be most important in a particular circumstance, such as motivational differences between teams or the impact of national-level political considerations on a local election (Kerr and Tindale 2011). Where a task lies on the intellective–judgmental spectrum affects the relative importance of social influence effects such as minority or majority influence. Rather than simply categorizing a context as intellective or judgmental and choosing a model accordingly, it would be preferable to define a parameter that gauges the balance between intellective and judgmental factors and therefore the weight of the dynamical mechanisms at play in a given context. Experiments could then test whether models integrated using the parameter could successfully make predictions for various tasks along the intellective–judgmental spectrum. As an illustration, one might conjecture that with respect to group polarization, the frame-induced mechanism might best suit forecasting problems, whereas political ideologies, falling further on the judgmental side, might best be modeled by extremist tilting. Policy-making might fall in between the two, and a parameter reflecting that balance could then be used to weight relative strengths of the frame-induced and extremist-tilting mechanisms.

A potential hazard of orienting experiments toward model testing is that experiments may be treated as primarily data-fitting exercises in which researchers test a raft of different models driven more by the various mathematical or simulation possibilities rather than by theory. This tendency will lead to models that are narrow in scope and not readily generalized to new circumstances. The SDS literature suffered from this tendency, as mentioned above, and never produced a compelling account of group polarization. However, the increasing prevalence of requirements among journals to make data sets available may help counter the lack of convergence caused by the tendency to seek and emphasize the best-fitting models for experiments in isolation. A new model, which best accounts for the results of a particular experiment, can now also be tested against data from previous experiments. The growth of a norm toward testing models against new and old data will encourage the development of more general models.

Initial successes in model development and testing will eventually lead to the emergence of a self-sustaining research community dedicated to the integration of modeling and experiment (an example of a new and virtuous epistemic culture as described in Chapter 2). In the short term, a comprehensive program aimed at providing experimental data to test and develop a range of models would help generate the nucleus of such a community as well as advancing social influence research itself. The goal of the program would be to develop general models that integrate different social influence phenomena over a range of contexts rather than the current practice that investigates behaviors in divergent research streams. To effect such synthesis, the program could unfold in phases in which experiments and models initially focus on relatively narrow phenomena with later phases becoming successively more integrative. This would encourage the development of more general and robust models and counter the tendency toward one-off model fitting. Follow-on research could test the models on real-world contexts of interest. Ideally, experiments would be conducted by separate teams of researchers who would then share the data with modeling teams. However, as there is little tradition in social psychology (and social science more broadly) of publishing experimental results without at least some theoretical embellishment, experimental teams could be allowed to develop their own models or included as coauthors on initial publications using their data.

While the goal of quantitatively testing computational models of social influence is challenging, the rewards for doing so would be high. On a scientific level, it would make the study of social influence more synthetic and cumulative. The bar would be raised for evaluating competing theories: a theory implemented in terms of a model that provided a quantitative account of experimental results would be preferable to one that only provided a qualitative account. In addition, the present practice of testing hypotheses experimentally based on

the coarse graining of variables results in nominal categorizations that can be difficult to extend to more general conditions and synthesize when competing effects and nonlinear interactions are present. Standard group polarization theory, for example, only applies when all group members have common inclinations. Although a more relaxed condition of *most* group members is often stated, that is too vague to enable prediction in situations where groups have members with opposing preferences. Effectively, the reference point is a wall beyond which standard polarization theory is silent. By allowing for gradations of effects, computational models are less constrained by such ambiguous categorizations; for instance, the rhetorical reference point in the ASC model is simply a parameter that affects the rhetorical distance between group members, a distance that can be calculated regardless of whether it spans the reference point or not. As a result, the ASC model can treat the combined F and U groups in Figure 13.1, while standard theory cannot. This freedom from dependence upon categorizations implies that models can be more readily extended to variable and parameter regimes not yet explored experimentally. In combination with the ability to probe the effects of nonlinear interactions via mathematical analysis and simulation, models can therefore be used to reveal novel, potentially counterintuitive behaviors not anticipated by qualitative theorizing.

Greater incentive to perform replication experiments would be another scientific benefit if experiments were to become more oriented toward computational models. Historically, there has been little incentive for social scientists to perform replications of previously reported effects and for journals to publish them. In the physical sciences, however, better measurements of model parameters, such as physical constants, are valued even if no new effect is reported, as such measurements improve the accuracy and precision of model predictions. Similarly, social influence experiments aimed at testing models would yield improved parameter estimation and so hold more value than merely replicating an effect. New experiments could repeat earlier ones but with higher resolution or an extended variable range, thereby enhancing the precision and robustness of parameter estimates. It is also possible that systematic deviations from model predictions could be observed pointing the way toward new theory and model development. A greater ability to publish such discrepant experimental data as valuable in its own right (without theoretical explanation) would therefore allow social scientists to learn more from data than is presently the case. Fundamental advances in physics have occurred because of the publication of experimental findings that ran counter to accepted theoretical models. A pivotal event in the genesis of quantum mechanics, for example, was the discovery of the photoelectric effect, a phenomenon at odds with classical physics, which Einstein eventually explained. The concentration on developing general models that minimize the number of free parameters will also discourage data dredging

practices in which researchers sift through a large number of covariates in order to find statistically significant, albeit likely spurious, relationships.

The approach outlined above is intended to develop a community within social influence research in which theory, modeling, and experiment proceed in a fashion similar to the physical sciences, albeit with reduced expectations of predictive power. It is not a call to end the traditional paradigm for the investigation of social influence. Although we have argued that the integrated approach will lead to new theories and discoveries, taking it further by demanding that novel theories be implemented formally before experimental testing would, on net, likely hamper the discovery of new behaviors, given the richness of social systems. A more desirable outcome would be for the model-oriented and traditional approaches to work in tandem. The standard testing of qualitative hypotheses could explore variables and effects not yet incorporated within quantitative models. This exploratory role would identify promising areas that could benefit from modeling and facilitate model development by narrowing the range of viable theoretical explanations. It is also possible that some behaviors will not be amenable to quantitative modeling and so remain in the province of qualitative theory in which modeling continues to play its more usual historical role in support of hypothesis generation.

Conclusion

Much of the recent surge of activity in computational social science has revolved around the analysis of massive amounts of data available from naturally occurring activity on the Internet and social media involving large networks consisting of thousands or millions of individuals. However, such studies do not shed light on the small group context, which is central to decision-making in leadership groups as well as political attitude change among ordinary citizens. Accordingly, the agenda put forth here emphasizes experiments with human subjects. Given their ability to control conditions, experimental studies can more directly test opinion models than can data from online networks or other observational sources. Network topology and initial opinion distributions can be controlled, the latter enabling testing of the core objective of modeling how opinions change from their initial values, rather than predicting final distributions on the basis of assumed initial conditions. Moreover, experimental results can provide a sounder basis for application of opinion network modeling to large systems. Such applications typically use models based on dyadic or other local interactions, and if those models cannot predict the results of small group experiments or be derived from

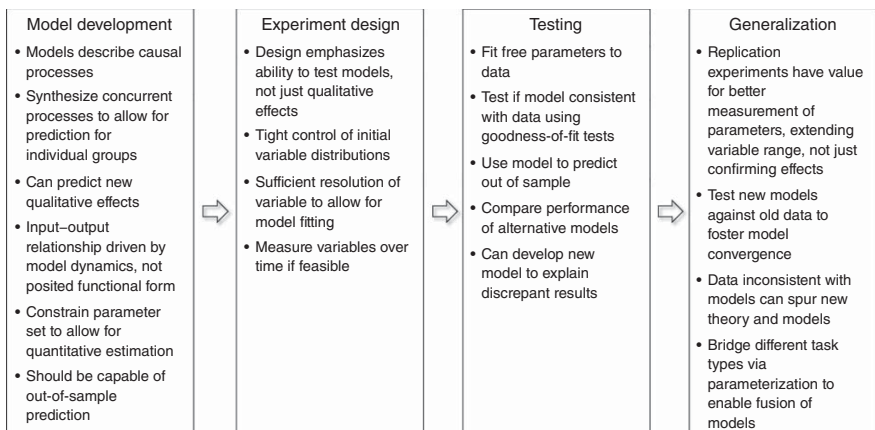


Figure 13.4 Overview of integrated modeling-experiment approach.

approximations of models that can, then little rationale exists for their use on large systems.

Figure 13.4 summarizes the integrated modeling and experimental approach proposed in this chapter. The goal of quantitatively testing computational models of social influence is no doubt an ambitious one. The approach advocated in this chapter centers upon the conduct of experiments explicitly designed to test the quantitative predictions of models rather than the standard experimental paradigm of testing qualitative hypotheses. Its aim is the development of models that can account for a range of phenomena and experimental results. Elements of this approach include: exercising discipline and discrimination with respect to model parameters, conducting goodness-of-fit tests, more highly resolved initial variable conditions, more deliberate control of initial opinion distributions, measuring opinions or other variables over time, greater use of out-of-sample prediction, testing models on new and old data to foster model convergence not proliferation, and parameterizing the nature of group tasks along a spectrum rather than ambiguously assigning them to nominal categories such as intellectual or judgmental.

While experimentation with human subjects is much more expensive and laborious than modeling and simulation, a greater emphasis on their integration will enhance both the influence of computational social science and the science of social influence. A major advantage of this integrated approach is an improved ability to synthesize different effects. Since opinion network models make predictions for specific groups, they must take more serious account of effects concurrent to the one under study, which otherwise might be assumed to wash out in a population of groups. Models can synthesize multiple effects more readily than combining different, often ambiguous, categorizations of conditions. The bar will be raised for the evaluation of rival theories with higher precedence given to theories whose associated models are in quantitative accord with experiment. Stronger incentive to conduct experiments for the purpose of providing better measurement or expanding the range of model variables – not just to test hypothesized relationships or competing theories – will be fostered under this approach. Greater replicability will ensue as will the ability to publish anomalous findings, thereby spurring new theory and model development. Ultimately, on an applications level, the integration of quantitative model testing and experiment will raise the confidence and scope with which models can be applied to natural situations for purposes of both prediction and designing interventions to shape outcomes.

Acknowledgments

This research was supported by the Office of Naval Research under grants N00014-15-1-2549 and N00014-16-1-2919.

References

- Becker, J., Brackbill, D., and Centola, D. (2017). Network dynamics of social influence in the wisdom of crowds. *Proceedings of the National Academy of Sciences* 114 (26): E5070–E5076. <https://doi.org/10.1073/pnas.1615978114>.
- Brown, R. (1986). *Social Psychology*, 2e. New York: Free Press.
- Burghardt, K., Rand, W., and Girvan, M. (2016). Competing opinions and stubbornness: connecting models to data. *Physical Review E* 93 (3): 032305.
- Castellano, C., Fortunato, S., and Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics* 81 (2): 591–646.
- Chong, D. and Druckman, J.N. (2007). Framing theory. *Annual Review of Political Science* 10 (1): 103–126. <https://doi.org/10.1146/annurev.polisci.10.072805.103054>.
- Crandall, D., Cosley, D., Huttenlocher, D. et al. (2008). Feedback effects between similarity and social influence in online communities. Paper presented at the Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, Las Vegas, Nevada, USA.
- Deffuant, G., Amblard, F., Weisbuch, G., and Faure, T. (2002). How can extremism prevail? A study based on the relative agreement interaction model. *Journal of Artificial Societies and Social Simulation* 5 (4).
- DeGroot, M.H. (1974). Reaching a consensus. *Journal of the American Statistical Association* 69 (345): 118–121. <https://doi.org/10.2307/2285509>.
- Devine, D.J. (2012). *Jury Decision Making: The State of the Science*. New York: NYU Press.
- Düring, B., Markowich, P., Pietschmann, J.-F., and Wolfram, M.-T. (2009). Boltzmann and Fokker–Planck equations modelling opinion formation in the presence of strong leaders. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science* 465 (2112): 3687–3708. <https://doi.org/10.1098/rspa.2009.0239>.
- Eagly, A.H. and Chaiken, S. (1993). *The Psychology of Attitudes*. Fort Worth, TX: Harcourt College Publishers.
- Fortunato, S. and Castellano, C. (2007). Scaling and universality in proportional elections. *Physical Review Letters* 99 (13): 138701.
- Friedkin, N.E. (1999). Choice shift and group polarization. *American Sociological Review* 64 (6): 856–875.
- Friedkin, N.E. (2015). The problem of social control and coordination of complex systems in sociology: a look at the community cleavage problem. *IEEE Control Systems* 35 (3): 40–51. <https://doi.org/10.1109/MCS.2015.2406655>.
- Friedkin, N.E. and Johnsen, E.C. (2011). *Social Influence Network Theory: A Sociological Examination of Small Group Dynamics*. Cambridge, UK: Cambridge University Press.
- Gabbay, M. (2007). The effects of nonlinear interactions and network structure in small group opinion dynamics. *Physica A* 378: 118–126.

- Gabbay, M. (2014). Data processing for applications of dynamics-based models to forecasting. In: *Sociocultural Behavior Sensemaking: State of the Art in Understanding the Operational Environment* (ed. J. Egeth, G. Klein and D. Schmorrow), 245–268. McLean, VA: The MITRE Corporation.
- Gabbay, M. (forthcoming). Opinion Network Modeling and Experiment. In: *Proceedings of the 5th International Conference on Theory and Applications in Nonlinear Dynamics* (ed. V. In, P. Longhini and A. Palacios). Springer.
- Gabbay, M., Kelly, Z., Reedy, J., and Gastil, J. (2018). Frame-induced group polarization in small discussion networks. *Social Psychology Quarterly* 81 (3): 248–271. <https://doi.org/10.1177/0190272518778784>.
- Hinsz, V.B. (1999). Group decision making with responses of a quantitative nature: the theory of social decision schemes for quantities. *Organizational Behavior and Human Decision Processes* 80 (1): 28–49. <http://dx.doi.org/10.1006/obhd.1999.2853>.
- Isenberg, D.J. (1986). Group polarization: a critical review and meta-analysis. *Journal of Personality and Social Psychology* 50 (6): 1141–1151. <http://dx.doi.org/10.1037/0022-3514.50.6.1141>.
- Keating, J., Van Boven, L., and Judd, C.M. (2016). Partisan underestimation of the polarizing influence of group discussion. *Journal of Experimental Social Psychology* 65: 52–58. <http://dx.doi.org/10.1016/j.jesp.2016.03.002>.
- Kerr, N.L. and Tindale, R.S. (2004). Group performance and decision making. *Annual Review of Psychology* 55 (1): 623–655.
- Kerr, N.L. and Tindale, R.S. (2011). Group-based forecasting?: a social psychological analysis. *International Journal of Forecasting* 27 (1): 14–40. <http://dx.doi.org/10.1016/j.ijforecast.2010.02.001>.
- Laughlin, P.R. and Ellis, A.L. (1986). Demonstrability and social combination processes on mathematical intellectual tasks. *Journal of Experimental Social Psychology* 22 (3): 177–189. [https://doi.org/10.1016/0022-1031\(86\)90022-3](https://doi.org/10.1016/0022-1031(86)90022-3).
- Lorenz, J. (2007). Continuous opinion dynamics under bounded confidence: a survey. *International Journal of Modern Physics C: Computational Physics and Physical Computation* 18 (12): 1819–1838.
- Mäs, M. and Flache, A. (2013). Differentiation without distancing. Explaining bi-polarization of opinions without negative influence. *PLoS One* 8 (11): e74516. <https://doi.org/10.1371/journal.pone.0074516>.
- Moussaïd, M., Kämmer, J.E., Analytis, P.P., and Neth, H. (2013). Social influence and the collective dynamics of opinion formation. *PLoS One* 8 (11): e78433. <https://doi.org/10.1371/journal.pone.0078433>.
- Moussaïd, M., Brighton, H., and Gaissmaier, W. (2015). The amplification of risk in experimental diffusion chains. *Proceedings of the National Academy of Sciences* 112 (18): 5631–5636. <https://doi.org/10.1073/pnas.1421883112>.

- Myers, D.G. (1982). Polarizing effects of social interaction. In: *Group Decision Making* (ed. H. Brandstatter, J.H. Davis and G. Stocker-Kreichauer). London: Academic Press.
- Myers, D.G. and Bishop, G.D. (1970). Discussion effects on racial attitudes. *Science* 169 (3947): 778–779. <https://doi.org/10.2307/1729790>.
- Olfati-Saber, R., Fax, J.A., and Murray, R.M. (2007). Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE* 95 (1): 215–233. <https://doi.org/10.1109/JPROC.2006.887293>.
- Pleskac, T.J., Diederich, A., and Wallsten, T.S. (2015). Models of decision making under risk and uncertainty. In: *The Oxford Handbook of Computational and Mathematical Psychology* (ed. J.R. Busemeyer, Z. Wang, J.T. Townsend and A. Eidels), 209–231. Oxford: Oxford University Press.
- Proskurnikov, A.V. and Tempo, R. (2017). A tutorial on modeling and analysis of dynamic social networks. Part I. *Annual Reviews in Control* 43 (Supplement C): 65–79. <https://doi.org/10.1016/j.arcontrol.2017.03.002>.
- Schkade, D., Sunstein, C.R., and Kahneman, D. (2000). Deliberating about dollars: the severity shift. *Columbia Law Review* 100 (4): 1139–1175. <https://doi.org/10.2307/1123539>.
- Schkade, D., Sunstein, C.R., and Hastie, R. (2010). When deliberation produces extremism. *Critical Review* 22 (2–3): 227–252. <https://doi.org/10.1080/08913811.2010.508634>.
- Schulz-Hardt, S. and Mojzisch, A. (2012). How to achieve synergy in group decision making: lessons to be learned from the hidden profile paradigm. *European Review of Social Psychology* 23 (1): 305–343. <https://doi.org/10.1080/10463283.2012.744440>.
- Sunstein, C.R. (2002). The law of group polarization. *Journal of Political Philosophy* 10 (2): 175–195.
- Török, J., Iñiguez, G., Yasserli, T. et al. (2013). Opinions, conflicts, and consensus: modeling social dynamics in a collaborative environment. *Physical Review Letters* 110 (8): 088701.
- Zaleska, M. (1982). The stability of extreme and moderate responses in different situations. In: *Group Decision Making* (ed. H. Brandstatter, J.H. Davis and G. Stocker-Kreichauer). London: Academic Press.

Appendix 4

Community Detectability and Structural Balance Dynamics in Signed Networks

Community detectability and structural balance dynamics in signed networks

Megan Morrison ^{*}*Department of Applied Mathematics, University of Washington, Washington 98115, USA*Michael Gabbay [†]*Applied Physics Laboratory, University of Washington, Washington 98115, USA*

(Received 17 December 2019; revised 14 May 2020; accepted 8 June 2020; published 6 July 2020)

We investigate signed networks with community structure with respect to their spectra and their evolution under a dynamical model of structural balance, a prominent theory of signed social networks. The spectrum of the adjacency matrix generated by a stochastic block model with two equal-size communities shows detectability transitions in which the community structure becomes manifest when its signal eigenvalue appears outside the main spectral band. The spectrum also exhibits “sociality” transitions involving the homogeneous structure representing the average tie value. We derive expressions for the eigenvalues associated with the community and homogeneous structure as well as the transition boundaries, all in good agreement with numerical results. Using the stochastically generated networks as initial conditions for a simple model of structural balance dynamics yields three outcome regimes: two hostile factions that correspond with the initial communities, two hostile factions uncorrelated with those communities, and a single harmonious faction of all nodes. The detectability transition predicts the boundary between the assortative and mixed two-faction states and the sociality transition predicts that between the mixed and harmonious states. Our results may yield insight into the dynamics of cooperation and conflict among actors with distinct social identities.

DOI: [10.1103/PhysRevE.102.012304](https://doi.org/10.1103/PhysRevE.102.012304)

I. INTRODUCTION

Most research in network science has focused on networks that allow only positive ties. In signed networks, however, ties can take on negative values as well. In social systems, positive ties signify friendly or cooperative relationships between the individual or collective actors represented by the nodes whereas negative ties signify hostile or conflictual relationships between nodes. As examples, signed social networks have been used to represent interpersonal sentiments among students [1], supportive or critical references among opinion makers [2], relationships in online social networks [3], and alliances and military clashes among nations [4,5].

In this paper, we address community structure in signed networks and its implications for dynamics governed by structural balance, a theory commonly invoked in treatments of signed networks in social systems. In unsigned networks, community structure refers to the presence of clusters within networks characterized by relatively dense intracluster ties and sparse intercluster ties. A rich set of techniques have been developed to detect communities in unsigned networks [6]. Of particular relevance here, spectral analysis has proven to be a highly valuable tool for probing community structure [7,8]. For signed networks, the notion of community can be extended to accommodate negative ties by reversing the crite-

ria for positive ties—there should be relatively sparse negative ties within communities and denser ties between them. At the present, however, the literature on community detection in signed networks is itself rather sparse in comparison with unsigned networks (e.g., Refs. [2,9–11]).

An important phenomenon of community structure in unsigned networks is that of community detectability [12–17]. Here, community structure can be present—in the sense that the tie-generating probabilities in a stochastic block model indeed favor ingroup over outgroup ties—but it is too weak to typically be discerned by analysis of the generated network. For large networks, a phase transition characterizes the passage from undetectable to detectable structure.

We show that detectability transitions also occur in signed networks. We generate our networks using a stochastic block model for two communities in an unweighted and undirected signed network (Sec. II). Examples of simulated networks with community structure that is detectable and undetectable are shown on the left in Figs. 1(a) and 1(b). We describe the transitions observed in the spectra of simulated networks in which outlying eigenvalues corresponding to meaningful signals merge with the main spectral band corresponding to noise (Sec. III). Two sets of spectral transitions are found: one corresponds to the detectability transition involving the two-community structure, while the other affects the ability to observe an overall tendency toward positive or negative tie formation, which we refer to as sociality transitions. Figure 1(c) shows an example of a network with a markedly positive average tie value generated in the regime in which an overall prosocial tendency can be reliably discerned.

^{*}mmtree@uw.edu[†]gabbay@uw.edu

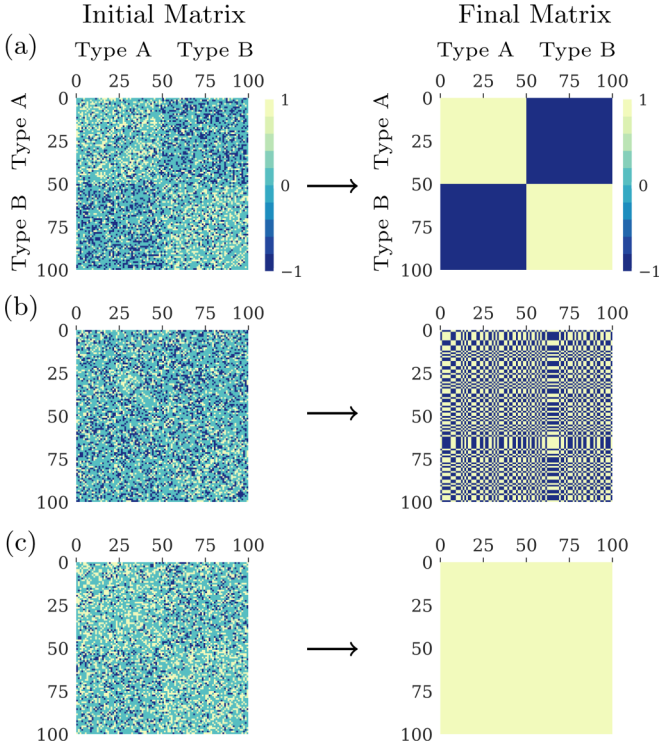


FIG. 1. Evolution of networks with initial community structure under structural balance dynamics. (a) Moderate initial structuring by group identity leads to a completely connected network consisting of two factions sorted by identity. (b) Weak initial structure leads to two factions of mixed identities. (c) Strong initial positivity in the network leads to a single harmonious faction. Networks represented as adjacency matrices with $\pm 1, 0$ tie values indicated by color. Initial networks generated by stochastic block model, Eqs. (2)–(4) with parameters $N = 100$ and $d_{\text{in}}, d_{\text{out}} = 0.4$ in all networks. $p_{\text{out}}^+ = 0.3$ and $p_{\text{in}}^+ = 0.7$ for (a), $p_{\text{out}}^+ = 0.4$ and $p_{\text{in}}^+ = 0.5$ for (b), and $p_{\text{out}}^+ = 0.6$ and $p_{\text{in}}^+ = 0.8$ for (c). Final networks represent the connectivity signs to which Eq. (1) converges (see Sec. VI).

We analytically calculate both the key eigenvalues and the transition conditions for large networks. In the main text, we use perturbation analysis to derive expressions for the signal eigenvalues (Sec. IV), which are then used to obtain the transition conditions by their equation with the main band edge eigenvalues (Sec. V), these edge eigenvalues being found using random matrix theory (Appendix A). We also present an alternative to our perturbation treatment that derives the signal eigenvalues on the basis of random matrix theory, in keeping with previous treatments of detectability (Appendix B) [13,14].

The spectral transitions have important implications for the outcomes of structural balance dynamics for networks possessing initial community structure. Structural balance theory, which postulates that triads with one or three negative edges will not endure, can be implemented as a deterministic, continuous time dynamical system (Sec. VI),

$$\frac{dY_{ij}}{dt} = \sum_{k=1}^N Y_{ik}Y_{kj}, \quad (1)$$

where t is time and N is the number of nodes [18,19]. This system evolves the connectivity Y_{ij} between nodes i and j as a function of their relationships with mutual neighbors: the product $Y_{ik}Y_{kj}$ increases their connectivity when they share a common inclination, positive or negative, toward k but decreases it if their inclinations are oppositely signed. This dynamic promotes balanced triads and eradicates unbalanced triads in the network. The model evolves into a fully connected network where either: (1) there are two hostile factions with only positive ties within each and only negative ties between them; or (2) all nodes are positively connected in a single harmonious faction. In either case, the final state is determined by the leading eigenvector of the initial network.

The driving role played by the leading eigenvector of the initial network in the structural balance evolution gives rise to a dynamical manifestation of the detectability transition when the leading eigenvector also carries the community structure signal. For the two-faction outcome, if the leading eigenvector corresponds to the two identity types in the stochastic block model, then the final factions will perfectly align with these identities as shown in Fig. 1(a). However, if the leading eigenvector is merely the edge of the main noise band, as occurs for weak initial structure below the detectability transition, then the composition of the final factions will not align with the identity types as seen in Fig. 1(b). An analogous transition to the single-faction outcome is generated by the sociality transition as seen in Fig. 1(c). Solutions of the structural balance model starting from networks randomly generated by the stochastic block model do indeed show sharp transitions between behavioral regimes whose boundaries agree with analytical predictions based on the detectability and sociality transitions (Sec. VII).

We discuss the potential implications of these results for conflict dynamics among actors with different identity types due to, for instance, ethnicity, religion, or ideology (Sec. VIII). In particular, conflicts such as civil wars may take on a binary nature. If the system starts out with weak identity-driven structure, then it will not be expected to polarize on the basis of identity. But complete identity polarization results even when initial affinities and animosities between identity types are fairly mild and even though identity itself plays no role in the micro-level conflict dynamics.

II. GENERATING AND REPRESENTING COMMUNITY STRUCTURE

Communities in an unsigned network are characterized by relatively dense within-community ties and sparse ties between communities. Community detection algorithms seek to discover these communities given an observed network [6,7,20]. Stochastic block models, which generate random networks with community structure by setting tie probabilities within and between blocks of nodes, have been used to investigate the behavior of community detection algorithms [21]. In this section, we describe the stochastic block model we use to generate our signed networks, the characterization of community structure via assortativity, and decomposition of the generated networks in terms of the eigenvectors of the average adjacency matrix and a random matrix.

A. Stochastic block model

Our construction starts with an undirected network of N nodes consisting of two identity groups A and B of equal size $N/2$, where $N \gg 1$. The A group nodes are indexed from 1 to $N/2$ and the B group from $N/2 + 1$ to N . $\mathbf{A}$ is the signed adjacency matrix where A_{ij} is the tie value between node i and node j , which can take on values of $\{1, -1, 0\}$ with 0 signifying the absence of a tie. As the network is undirected, the adjacency matrix is symmetric, $A_{ij} = A_{ji}$. The probability that a tie, positive or negative, will form between any given ingroup (A with A, B with B) node pair is d_{in} . Similarly, the tie formation probability between outgroup (A with B) node pairs is d_{out} . These tie formation probabilities are equivalent to the expected ingroup and outgroup tie densities and their average yields the expected tie density for the total network, $d = (d_{\text{in}} + d_{\text{out}})/2$. Given the presence of a tie between ingroup members, the conditional probability that it is positive is p_{in}^+ and that it is negative is $p_{\text{in}}^- = 1 - p_{\text{in}}^+$. Similarly, the positive and negative tie conditional probabilities between outgroup nodes are written p_{out}^+ and $p_{\text{out}}^- = 1 - p_{\text{out}}^+$. For brevity, we refer to p_{in}^+ and p_{out}^+ as the ingroup and outgroup *affinities* and p_{in}^- and p_{out}^- as the in and outgroup *animosities*.

The adjacency matrix can be written in terms of the following block structure:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{AA} & \mathbf{A}_{AB} \\ \mathbf{A}_{BA} & \mathbf{A}_{BB} \end{bmatrix}, \quad (2)$$

where each block is a random $N/2 \times N/2$ matrix. The diagonal blocks represent AA or BB ties, whose elements are set using the following probability distribution for the ingroup random variable A_{in} :

$$\mathbb{P}(A_{\text{in}} = k) = \begin{cases} d_{\text{in}} p_{\text{in}}^+, & k = 1, \\ d_{\text{in}}(1 - p_{\text{in}}^+), & k = -1, \\ 1 - d_{\text{in}}, & k = 0. \end{cases} \quad (3)$$

Since $\mathbf{A}$ is symmetric, there are $(N/2)(N/2 + 1)$ independent, identically distributed (i.i.d.) ingroup ties.

The off-diagonal blocks, corresponding to AB or BA ties, are transposes of each other resulting in $N^2/4$ i.i.d. outgroup ties, which are drawn according to the random variable A_{out} :

$$\mathbb{P}(A_{\text{out}} = k) = \begin{cases} d_{\text{out}} p_{\text{out}}^+, & k = 1, \\ d_{\text{out}}(1 - p_{\text{out}}^+), & k = -1, \\ 1 - d_{\text{out}}, & k = 0. \end{cases} \quad (4)$$

Equations (2)–(4) define the stochastic block model used to generate matrices with more or less community structure as seen on the left in Fig. 1 and for all the numerically generated spectra shown in this paper. Note that this model allows nonzero self-ties, unlike in many empirical networks, but this is a standard approximation that facilitates analytical treatment [13,14,22]. For large networks, the contribution of the N diagonal elements is negligible in comparison with that of the order N^2 off-diagonal elements. As we note below, the effect of removing self-ties on the average of $\mathbf{A}$ is to shift the eigenvalues by a constant that is independent of N . In addition, the model of structural balance dynamics, Eq. (1), allows for self-ties.

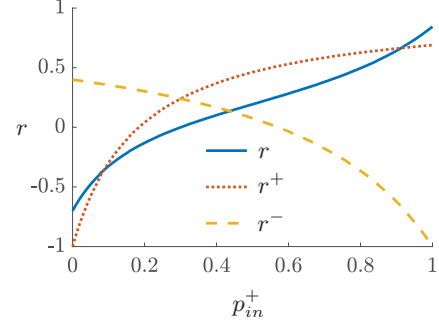


FIG. 2. Example of assortativity coefficients as a function of ingroup affinity. Values calculated using expected tie numbers. Parameters are $d_{\text{in}} = 0.5$, $d_{\text{out}} = 0.3$, $p_{\text{out}}^- = 0.7$, and $N = 100$.

Networks that are too sparse become disconnected and the structural balance dynamics of the isolated subgroups will evolve independently of each other as opposed to the holistic evolution we see when the network forms one connected graph. An Erdős-Rényi graph will very likely form a single connected graph if $p > \ln(N)/N$ [17,23]. Similarly, our stochastic block model matrices with two communities and density probabilities $d_{\text{in}} = a/N$ and $d_{\text{out}} = b/N$ will have a single giant component if $(a + b)/2 > 1$, and will very likely form a single connected graph if $d_{\text{in}} = a \ln(N)/N$, $d_{\text{out}} = b \ln(N)/N$ and $(a + b)/2 > 1$ [17].

B. Assortativity

Assortativity refers to the tendency for nodes of the same type to be more strongly connected than nodes of different types. We extend the standard definition of the assortativity coefficient for discrete node types [24] to our signed network case by calculating separate coefficients for the positive and negative tie networks and then essentially differencing them. We will use the signed network assortativity coefficient to characterize the regimes of the structural balance dynamics in Sec. VII.

First, considering the adjacency matrix of positive ties only, we let e_{ij}^+ denote the fraction of all positive ties that connect a node of type i to one of type j where $i, j \in \{A, B\}$. The assortativity coefficient r^+ for the network of positive ties, whose adjacency matrix elements are 1 if $A_{ij} > 0$ and zero otherwise, is then

$$r^+ = \frac{\sum_i e_{ii}^+ - \sum_i (a_i^+)^2}{1 - \sum_i (a_i^+)^2}, \quad (5)$$

where $a_i^+ = \sum_j e_{ij}^+$. The assortativity coefficient can range between -1 and 1 . A network containing only ingroup (AA or BB) positive ties with no outgroup (AB, BA) ties is completely assortative, $r^+ = 1$, which in the social network context implies that people only cooperate with members of the same group. A network containing only outgroup ties is completely disassortative, $r^+ = -1$, implying cooperation across the two groups but not within them. We see this state in the example shown in Fig. 2 in which $r^+ = -1$ when the ingroup affinity $p_{\text{in}}^+ = 0$. As the ingroup affinity increases, r^+ increases but does not reach one as there are still outgroup ties due to the nonzero value of the fixed outgroup affinity.

The assortativity for the network of negative ties, r^- , is defined analogously to Eq. (5). Whereas positive values of r^+ imply that ingroup relations are more friendly than outgroup relations, assortative mixing in the network of negative ties implies more hostility within groups than between them. Thus, in Fig. 2 we see that $r^- > 0$ when there is complete ingroup animosity ($p_{\text{in}}^- = 1$ corresponding to $p_{\text{in}}^+ = 0$) and $r^- = -1$ when there is no ingroup animosity and so no negative ties within groups.

Accordingly, as we want positive values of our overall signed network assortativity coefficient r to signify that ingroup interactions tend to be more amicable than outgroup ones, we average r^+ and $-r^-$, yielding

$$r = \frac{r^+ - r^-}{2}, \quad (6)$$

which can take on values between -1 and 1 . Figure 2 shows that r is negative for low ingroup affinity and positive for high ingroup affinity.

C. Adjacency matrix decomposition

This section presents a decomposition of the adjacency matrices generated by the stochastic block model into: (i) a signal component that results from the expected tie values generated by the ingroup and outgroup random variables, A_{in} and A_{out} ; and (ii) a noise component due to random deviations from the expected values. This decomposition will form the starting point for our calculation of the network eigenvalues in Sec. IV. We write $\mathbf{A}$ as the sum of the average matrix $\langle \mathbf{A} \rangle$ and a random deviation matrix $\mathbf{X}$:

$$\mathbf{A} = \langle \mathbf{A} \rangle + \mathbf{X}. \quad (7)$$

Given the block structure of Eq. (2), $\langle \mathbf{A} \rangle$ can be written as

$$\langle \mathbf{A} \rangle = \begin{bmatrix} \langle A_{AA} \rangle & \langle A_{AB} \rangle \\ \langle A_{BA} \rangle & \langle A_{BB} \rangle \end{bmatrix}, \quad (8)$$

where each element of $\langle A_{AA} \rangle$ and $\langle A_{BB} \rangle$ is equal to $\langle A_{\text{in}} \rangle$ and each element of $\langle A_{AB} \rangle$ and $\langle A_{BA} \rangle$ is equal to $\langle A_{\text{out}} \rangle$. From Eqs. (3) and (4), we have

$$\langle A_{\text{in}} \rangle = d_{\text{in}}(2p_{\text{in}}^+ - 1), \quad (9)$$

$$\langle A_{\text{out}} \rangle = d_{\text{out}}(2p_{\text{out}}^+ - 1). \quad (10)$$

We define a couple of useful linear combinations of the in and outgroup expected tie values. We denote by μ the average over all the elements in $\langle \mathbf{A} \rangle$,

$$\mu = \frac{\langle A_{\text{in}} \rangle + \langle A_{\text{out}} \rangle}{2}, \quad (11)$$

and we denote by ν the half-difference between the in and outgroup expected tie values,

$$\nu = \frac{\langle A_{\text{in}} \rangle - \langle A_{\text{out}} \rangle}{2}. \quad (12)$$

Both μ and ν range from -1 to 1 . Noting that $\langle A_{\text{in}} \rangle = \mu + \nu$ and $\langle A_{\text{out}} \rangle = \mu - \nu$, these expressions allow us to express $\langle \mathbf{A} \rangle$ as a sum of two outer products,

$$\langle \mathbf{A} \rangle = \mu N \mathbf{u}_H \mathbf{u}_H^T + \nu N \mathbf{u}_C \mathbf{u}_C^T, \quad (13)$$

where $\mathbf{u}_H$ and $\mathbf{u}_C$ are orthonormal N -dimensional vectors: $\mathbf{u}_H = \frac{1}{\sqrt{N}}[1, 1, \dots, 1]^T$ and $\mathbf{u}_C = \frac{1}{\sqrt{N}}[1, \dots, 1, -1, \dots, -1]^T$, where the -1 's align with the B block node indices. In fact, $\mathbf{u}_H$ and $\mathbf{u}_C$ are readily seen to be the two eigenvectors of $\langle \mathbf{A} \rangle$ with respective eigenvalues μN and νN :

$$\langle \mathbf{A} \rangle \mathbf{u}_H = \mu N \mathbf{u}_H, \quad (14)$$

$$\langle \mathbf{A} \rangle \mathbf{u}_C = \nu N \mathbf{u}_C. \quad (15)$$

The term containing $\mathbf{u}_H$ in Eq. (13) generates a homogeneous $N \times N$ matrix whose elements are all equal to μ , the global average tie value. Hence, we refer to $\mathbf{u}_H$ as the *homogeneous* eigenvector. The term containing $\mathbf{u}_C$ generates a matrix whose diagonal block elements are all equal to ν and whose off-diagonal block elements are $-\nu$ and so corresponds to the structure of ingroup and outgroup tie differences. Accordingly, $\mathbf{u}_C$ generates the community structure and we refer to it as the *contrast* eigenvector.

The homogeneous and contrast eigenvectors are *signal* eigenvectors whose ability to be distinguished from the noise generated by $\mathbf{X}$ has important implications for community detectability and structural balance dynamics. From this perspective, μ and ν can be regarded as natural parameters for the signal structure in the network and could be used in place of two of the parameters in the stochastic block model, for instance, the ingroup and outgroup affinities. Doing so is less intuitive from a simulation viewpoint, however.

While $\langle \mathbf{A} \rangle$ is rank 2, in general, if either μ or ν equals zero, then $\langle \mathbf{A} \rangle$ becomes a rank 1 matrix composed of either the homogeneous eigenvector or the contrast eigenvector. If $p_{\text{out}}^+ = 1/2$, then $\langle A_{\text{out}} \rangle = 0$ and $\mu = \nu$ and so the two eigenvalues of $\langle \mathbf{A} \rangle$ are degenerate, as is the case when $d_{\text{out}} = 0$ and the two blocks are disconnected from each other. However, taking $d_{\text{in}} = 0$ yields a bipartite network between the A and B blocks in which $\mu = \langle A_{\text{out}} \rangle / 2 = -\nu$, and so the homogeneous and contrast eigenvalues are equal and opposite. If μ and ν are both zero, then $\langle \mathbf{A} \rangle$ vanishes and $\mathbf{A}$ reduces to the noise matrix $\mathbf{X}$.

To remove self-ties from $\langle \mathbf{A} \rangle$, one can subtract $\langle A_{\text{in}} \rangle \mathbf{I}$ from Eq. (13), where $\mathbf{I}$ is the identity matrix. This shifts the signal eigenvalues by $-\langle A_{\text{in}} \rangle = -(\mu + \nu)$.

The noise matrix $\mathbf{X}$ is a symmetric matrix that can be written in the block form,

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{AA} & \mathbf{X}_{AB} \\ \mathbf{X}_{BA} & \mathbf{X}_{BB} \end{bmatrix}. \quad (16)$$

Since $\mathbf{X} = \mathbf{A} - \langle \mathbf{A} \rangle$, the elements of the ingroup blocks $\mathbf{X}_{AA}$ and $\mathbf{X}_{BB}$ can assume values in $\{1 - \langle A_{\text{in}} \rangle, -1 - \langle A_{\text{in}} \rangle, -\langle A_{\text{in}} \rangle\}$ that are distributed according to the random variable X_{in} ,

$$\mathbb{P}(X_{\text{in}} = k) = \begin{cases} d_{\text{in}} p_{\text{in}}^+, & k = 1 - \langle A_{\text{in}} \rangle, \\ d_{\text{in}}(1 - p_{\text{in}}^+), & k = -1 - \langle A_{\text{in}} \rangle, \\ 1 - d_{\text{in}}, & k = -\langle A_{\text{in}} \rangle. \end{cases} \quad (17)$$

Likewise, the entries of the outgroup blocks $\mathbf{X}_{AB} = \mathbf{X}_{BA}^T$ are distributed like X_{out} ,

$$\mathbb{P}(X_{\text{out}} = k) = \begin{cases} d_{\text{out}} p_{\text{out}}^+, & k = 1 - \langle A_{\text{out}} \rangle, \\ d_{\text{out}}(1 - p_{\text{out}}^+), & k = -1 - \langle A_{\text{out}} \rangle, \\ 1 - d_{\text{out}}, & k = -\langle A_{\text{out}} \rangle. \end{cases} \quad (18)$$

All the elements of $\mathbf{X}$ have zero mean as $\langle X_{\text{in}} \rangle = \langle X_{\text{out}} \rangle = 0$. The variances of X_{in} and X_{out} are given by $\sigma_{\text{in}}^2 = \langle X_{\text{in}}^2 \rangle$ and $\sigma_{\text{out}}^2 = \langle X_{\text{out}}^2 \rangle$, which are written in terms of the stochastic block model parameters as

$$\sigma_{\text{in}}^2 = d_{\text{in}} - d_{\text{in}}^2(2p_{\text{in}}^+ - 1)^2, \quad (19)$$

$$\sigma_{\text{out}}^2 = d_{\text{out}} - d_{\text{out}}^2(2p_{\text{out}}^+ - 1)^2. \quad (20)$$

These variances will appear as their average,

$$\sigma^2 = \frac{\sigma_{\text{in}}^2 + \sigma_{\text{out}}^2}{2}, \quad (21)$$

in the noise-induced correction to the signal eigenvalues calculated below. The average variance can also be related to the parameters μ and ν as follows:

$$\sigma^2 = \frac{d_{\text{in}} + d_{\text{out}}}{2} - \frac{d_{\text{in}}^2(2p_{\text{in}}^+ - 1)^2 + d_{\text{out}}^2(2p_{\text{out}}^+ - 1)^2}{2}, \quad (22)$$

$$= \frac{d_{\text{in}} + d_{\text{out}}}{2} - \frac{1}{2}(\langle A_{\text{in}} \rangle^2 + \langle A_{\text{out}} \rangle^2), \quad (23)$$

$$= \frac{d_{\text{in}} + d_{\text{out}}}{2} - \mu^2 - \nu^2, \quad (24)$$

where we have used Eqs. (9) and (10) in the second line and $\langle A_{\text{in}} \rangle^2 + \langle A_{\text{out}} \rangle^2 = (\mu + \nu)^2 + (\mu - \nu)^2 = 2(\mu^2 + \nu^2)$ in the third.

III. SIGNAL EIGENVALUE TRANSITIONS

Spectral analysis has been used to address the number and detectability of communities in unsigned networks by considering the leading eigenvalues that reside outside the (approximately) continuous main spectral band due to its generation as a random graph [8,13,14,25]. For undirected networks, the adjacency matrix is symmetric and hence has a real spectrum. The number of detectable communities is equivalent to the number of positive eigenvalues that lie beyond the main spectral band. Nadakuditi and Newman [13] showed the existence of, and analytically calculated, a detectability transition in which the community structure, as generated by a stochastic block model with two communities, while still present becomes no longer detectable. Under assortative tie formation, this transition occurs once the second eigenvalue of the adjacency matrix, which carries the community structure information, merges with the main spectral band. Using random matrix theory, the authors derived expressions for both leading eigenvalues and the edge of the spectral band, thereby enabling the analytical determination of the transition dependence.

Similar to unsigned networks, Fig. 3 illustrates that the spectra of our signed networks consist of a continuous band of eigenvalues originating from $\mathbf{X}$, the variability or noise in the system, and signal eigenvalues originating from $\langle \mathbf{A} \rangle$, the structure in the system. The edges of the main spectral band, $\pm\gamma$, are derived in Appendix A,

$$\gamma = 2\sigma\sqrt{N}, \quad (25)$$

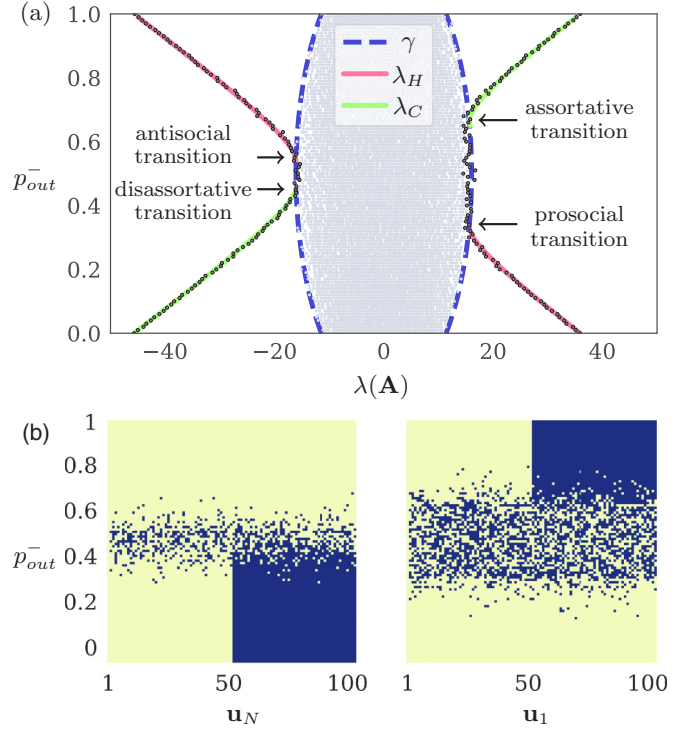


FIG. 3. (a) Eigenvalue spectrum of $\mathbf{A}$ as a function of out-group animosity p_{out}^- . The other parameters remain constant: $N = 100$, $p_{\text{in}}^+ = 0.4$, $d_{\text{in}} = 0.5$, $d_{\text{out}} = 0.8$. (b) Signs (yellow, positive; blue, negative) of the components of the last, $\mathbf{u}_N$, and first, $\mathbf{u}_1$, eigenvectors as functions of p_{out}^- . The theoretical curves for λ_C (solid green), λ_H (solid pink), and γ (dashed blue) are calculated from Eqs. (26), (27), and (A15), respectively.

while the average contrast and homogeneous signal eigenvalues, $\langle \lambda_C \rangle$ and $\langle \lambda_H \rangle$, are derived in Sec. IV and Appendix B,

$$\langle \lambda_C \rangle = \nu N + \frac{\sigma^2}{\nu}, \quad |\nu| \geq \frac{\sigma}{\sqrt{N}}, \quad (26)$$

$$\langle \lambda_H \rangle = \mu N + \frac{\sigma^2}{\mu}, \quad |\mu| \geq \frac{\sigma}{\sqrt{N}}. \quad (27)$$

The formulas for $\langle \lambda_C \rangle$ and $\langle \lambda_H \rangle$ consist of their respective eigenvalues from $\langle \mathbf{A} \rangle$ and a correction proportional to the average variance. The conditions $|\nu|, |\mu| \geq \sigma/\sqrt{N}$ imply that $|\langle \lambda_C \rangle|, |\langle \lambda_H \rangle| \geq |\gamma|$ so that Eqs. (26) and (27) are only valid when outside of the main spectral band. Note that because of the self-averaging behavior for large N , we can effectively drop the expectation brackets and take $\lambda_C \approx \langle \lambda_C \rangle$ and $\lambda_H \approx \langle \lambda_H \rangle$.

The successive horizontal slices in Fig. 3(a) correspond to the eigenvalues of single instances of $\mathbf{A}$ generated by the stochastic block model as the outgroup animosity is increased. There are four points at which the outlying eigenvalues merge with the main band. Considering first the upper right of the plot, the largest eigenvalue, λ_1 , is observed to detach from the main band for p_{out}^- greater than about 0.7. The right plot in Fig. 3(b), which depicts the signs of components of the first eigenvector $\mathbf{u}_1$, shows that $\mathbf{u}_1$ displays a two block structure for high p_{out}^- . Consequently, in this regime, the leading eigenvector corresponds to a perturbed version

of the contrast eigenvector, $\mathbf{u}_C$, of $\langle \mathbf{A} \rangle$. The point at which λ_1 emerges from the main band is then identified with the community detectability transition. Indeed, below this point $\mathbf{u}_1$ loses its block structure and rapidly takes on the appearance of random noise. We observe that in contrast to the analogous case in unsigned networks where the detectability transition involves the second eigenvalue of the adjacency matrix, here the two communities are no longer discernible when the leading eigenvalue merges with the main spectral band. We refer to the transition involving the merging of the contrast eigenvalue with the positive edge of the main band as the *assortative* transition because the communities are preferentially grouped by identity type. With an eye toward its dynamical significance when the contrast eigenvalue is leading, the two final factions produced by the structural balance dynamics are polarized by identity type after λ_1 emerges from the band.

The leading eigenvalue is observed to undergo another transition at $p_{\text{out}}^- \approx 0.3$. For lower outgroup animosities, $\mathbf{u}_1$ takes on a single block structure and so can be taken to result from a perturbation of the homogeneous eigenvector $\mathbf{u}_H$. This homogeneous structure disappears from $\mathbf{u}_1$ for p_{out}^- values above the transition. As the homogeneous eigenvalue carries information about the average tie value over all nodes, its emergence from the noise band can be considered a *sociality* transition. In particular, we refer to transitions that occur on the positive side of the noise band as *prosocial* transitions, in which a pattern of overall positive ties between nodes becomes apparent. The prosocial transition induces a transition in the structural balance dynamics from the two-faction equilibrium (not sorted by identity) to a single harmonious faction consisting of all nodes.

The lower left section of Fig. 3(a) shows the intersection of the last and least eigenvalue, λ_N , with the noise band at $p_{\text{out}}^- \approx 0.4$. For outgroup animosity values beneath this intersection, the last eigenvector $\mathbf{u}_N$ displays a two block structure as seen on the left plot of Fig. 3(b). However, although these blocks align with the A and B identity groups, the warmer outgroup than ingroup relations implies that the blocks are really disassortative “anti”-communities rather than assortative communities (prominent negative eigenvalues are also associated with disassortativity in unsigned networks [7]). Since this *disassortative* transition involves the least eigenvector, it has no significance with respect to the outcomes of the structural balance dynamics.

Finally, the other transition involving λ_N , seen in the upper left of Fig. 3(a), represents the emergence of the homogeneous eigenvalue and its corresponding single block structure from the noise band. It is a sociality transition and, in particular, an *antisocial* transition as it occurs on the negative side of the noise band signifying a conflictual relationship among nodes on average. The antisocial transition has no dynamical significance with respect to the structural balance dynamics.

Signal eigenvalues can occur on alternative sides of the spectral band as in Fig. 3(a) or on the same side as seen in Fig. 4(a). When λ_H and λ_C are on the same side of the spectral band, $p_{\text{out}}^- = 1/2$ is the point at which the signal eigenvalues cross one another since $\mu = \nu$. This affects which of the first two eigenvectors carries the community structure, but the structure itself remains apparent [Fig 4(b)]. However, the signal crossing does affect the balance dynamics,

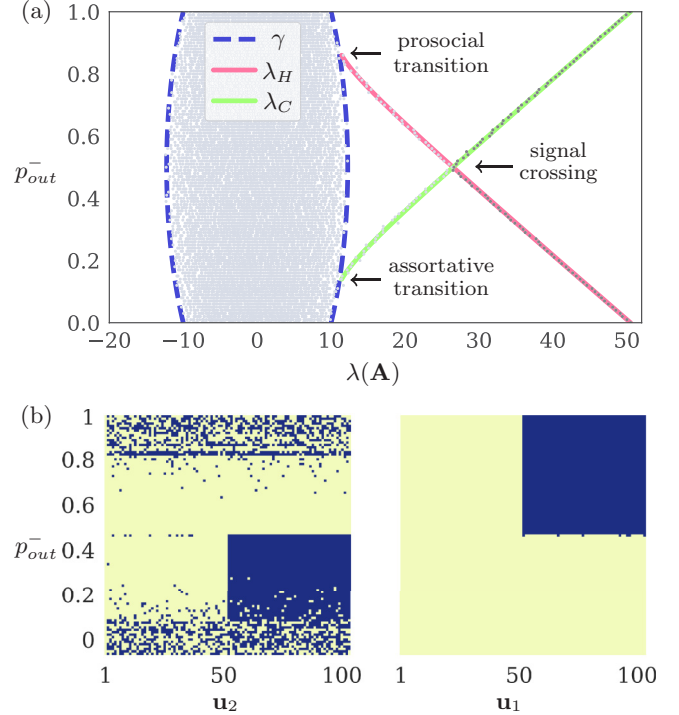


FIG. 4. (a) Spectrum of $\mathbf{A}$ as a function of p_{out}^- with constant parameters $N = 100$, $d_{\text{in}}, d_{\text{out}} = 0.5$, and $p_{\text{in}}^+ = 1$. (b) Signs (yellow positive, blue negative) of the components of the first and second eigenvectors, $\mathbf{u}_1$ and $\mathbf{u}_2$, as functions of p_{out}^- . The theoretical curves for λ_C (solid green), λ_H (solid pink), and γ (dashed blue) are calculated from Eqs. (26), (27), and (A15), respectively.

producing a transition between the harmonious and assortative outcomes.

IV. CALCULATION OF SIGNAL EIGENVALUES

In this section, we will derive formulas for the signal eigenvalues as a function of our stochastic block model parameters. We employ a perturbation treatment here but present an alternative derivation employing random matrix theory and complex analysis in Appendix B. Equation (7), which expresses the adjacency matrix $\mathbf{A}$ as the sum of its expected value $\langle \mathbf{A} \rangle$ and a matrix of random deviations $\mathbf{X}$, will form the starting point of our analysis. We consider $\langle \mathbf{A} \rangle$ as a given deterministic matrix with homogeneous and contrast eigenvalues and eigenvectors, Eqs. (14) and (15), that is subject to a perturbation from the independent noise matrix $\mathbf{X}$, which induces shifts to the signal eigenvalues and eigenvectors. A perturbation expansion to second order and the statistics of the noise matrix then yield the corrections to the signal eigenvalues. We note that in treating $\mathbf{X}$ as an independent perturbation to $\langle \mathbf{A} \rangle$, we temporarily suspend their linkage via the tie formation probabilities in the stochastic block model.

A. Perturbation expansion setup

We show the perturbation calculation for the case of the contrast eigenvalue. The homogeneous eigenvalue can be

obtained in precisely analogous fashion. The eigenvalue equation is

$$(\langle \mathbf{A} \rangle + \mathbf{X})\mathbf{v}_C = \lambda_C \mathbf{v}_C. \quad (28)$$

We write the perturbed eigenvector and eigenvalue up to second order as

$$\mathbf{v}_C = \mathbf{u}_C + \mathbf{u}^{(1)} + \mathbf{u}^{(2)}, \quad (29)$$

$$\lambda_C = \nu N + \lambda^{(1)} + \lambda^{(2)}, \quad (30)$$

where $\lambda^{(1)}$, $\lambda^{(2)}$, $\mathbf{u}^{(1)}$, $\mathbf{u}^{(2)}$ are the first- and second-order perturbations. As the unperturbed eigenvalue is $\mathcal{O}(\nu N)$, we divide Eq. (28) by N so that the zeroth-order equation is $\mathcal{O}(1)$. Then, using Eqs. (29) and (30), the eigenvalue equation becomes

$$\begin{aligned} & \left(\frac{\langle \mathbf{A} \rangle}{N} + \frac{\mathbf{X}}{N} \right) (\mathbf{u}_C + \mathbf{u}^{(1)} + \mathbf{u}^{(2)}) \\ &= \left(\nu + \frac{\lambda^{(1)}}{N} + \frac{\lambda^{(2)}}{N} \right) (\mathbf{u}_C + \mathbf{u}^{(1)} + \mathbf{u}^{(2)}). \end{aligned} \quad (31)$$

Before embarking upon our perturbation analysis, we specify the appropriate expansion parameter. To do so, we determine the orders of the $\langle \mathbf{A} \rangle$ and $\mathbf{X}$ matrices by evaluating their 2-norms. The 2-norm of each matrix is equivalent to its largest eigenvalue. Consequently, for the unperturbed matrix, $\|\langle \mathbf{A} \rangle\|_2$ is given by the larger of $N|\nu|$ or $N|\mu|$. Taking ν, μ to be $\mathcal{O}(1)$ therefore implies that $\langle \mathbf{A} \rangle$ is $\mathcal{O}(N)$. In Appendix A, using Wigner's semicircle law for the spectral density of a random matrix as well as matrix bounds, we determine that $\|\mathbf{X}\|_2$ is $\mathcal{O}(\sigma\sqrt{N})$. The ratio of the orders of $\mathbf{X}$ to $\langle \mathbf{A} \rangle$ is $\mathcal{O}(\sigma/\sqrt{N})$, and so successive orders in the perturbation series must diminish by a factor of $\sigma/\sqrt{N}$, which therefore serves as our expansion parameter. At a given N , the perturbation can be made arbitrarily small by letting σ go to zero. But in the large N regime, we need not constrain σ to be small.

Separating Eq. (31) out by expansion orders yields $\mathcal{O}(1)$

$$\frac{\langle \mathbf{A} \rangle}{N} \mathbf{u}_C = \nu \mathbf{u}_C, \quad (32)$$

$\mathcal{O}(\sigma/\sqrt{N})$

$$\frac{\langle \mathbf{A} \rangle}{N} \mathbf{u}^{(1)} + \frac{\mathbf{X}}{N} \mathbf{u}_C = \nu \mathbf{u}^{(1)} + \frac{\lambda^{(1)}}{N} \mathbf{u}_C, \quad (33)$$

$\mathcal{O}(\sigma^2/N)$

$$\frac{\langle \mathbf{A} \rangle}{N} \mathbf{u}^{(2)} + \frac{\mathbf{X}}{N} \mathbf{u}^{(1)} = \nu \mathbf{u}^{(2)} + \frac{\lambda^{(1)}}{N} \mathbf{u}^{(1)} + \frac{\lambda^{(2)}}{N} \mathbf{u}_C. \quad (34)$$

B. First-order treatment

To find the first-order eigenvalue perturbation, $\lambda^{(1)}$, we multiply both sides of Eq. (33) by $N\mathbf{u}_C^T$. Then using Eq. (13) and the orthonormality of $\mathbf{u}_C$ and $\mathbf{u}_H$ gives

$$\nu N \mathbf{u}_C^T \mathbf{u}^{(1)} + \mathbf{u}_C^T \mathbf{X} \mathbf{u}_C = \nu N \mathbf{u}_C^T \mathbf{u}^{(1)} + \lambda^{(1)}. \quad (35)$$

Solving for $\lambda^{(1)}$ yields

$$\begin{aligned} \lambda^{(1)} &= \mathbf{u}_C^T \mathbf{X} \mathbf{u}_C \\ &= \frac{1}{N} \left\{ \sum_{i,j=1}^{\frac{N}{2}} X_{ij} + \sum_{i,j=\frac{N}{2}+1}^N X_{ij} - 2 \sum_{i=1}^{\frac{N}{2}} \sum_{j=\frac{N}{2}+1}^N X_{ij} \right\}. \end{aligned} \quad (36)$$

The first two terms in the braces above sum ties in the ingroup blocks AA and BB , respectively, each tie distributed as X_{in} , and the third term corresponds to the AB and BA outgroup ties, distributed as X_{out} . As $\lambda^{(1)}$ is equal to the sum of zero-mean random ingroup and outgroup variables, its mean therefore vanishes,

$$\langle \lambda^{(1)} \rangle = 0. \quad (37)$$

Turning to the variance, each element within the outgroup sum has variance σ_{out}^2 , which becomes $4\sigma_{\text{out}}^2/N^2$ when the $2/N$ prefactor is included. The contribution to the variance from the $N^2/4$ outgroup variables is therefore σ_{out}^2 . Similarly, for the ingroup sums, the symmetry of X implies that, neglecting the diagonal, there are approximately a total of $N^2/4$ independent variables each with variance $4\sigma_{\text{in}}^2/N^2$ so that the ingroup variance contribution is σ_{in}^2 . Accordingly, the variance of $\lambda^{(1)}$ is

$$\text{Var}(\lambda^{(1)}) = \sigma_{\text{in}}^2 + \sigma_{\text{out}}^2 = 2\sigma^2. \quad (38)$$

To solve for $\mathbf{u}^{(1)}$, we write it as a vector decomposition and solve for the individual components,

$$\mathbf{u}^{(1)} = \mathbf{u}_{\parallel}^{(1)} + \mathbf{u}_{\perp}^{(1)}, \quad (39)$$

where $\mathbf{u}_{\parallel}^{(1)}$ is the component of $\mathbf{u}^{(1)}$ that is in the $\mathbf{u}_C, \mathbf{u}_H$ plane, and $\mathbf{u}_{\perp}^{(1)}$ is the component orthogonal to that plane. We find $\mathbf{u}_{\parallel}^{(1)}$ and $\mathbf{u}_{\perp}^{(1)}$ by multiplying both sides of the $\mathcal{O}(\sigma/\sqrt{N})$ equation, Eq. (33) by $\mathbf{u}_H^T$, the transpose of the homogeneous eigenvector of $\langle \mathbf{A} \rangle$, which gives

$$\mathbf{u}_H^T (\langle \mathbf{A} \rangle \mathbf{u}^{(1)} + \mathbf{X} \mathbf{u}_C) = \mathbf{u}_H^T (\nu N \mathbf{u}^{(1)} + \lambda^{(1)} \mathbf{u}_C), \quad (40)$$

which after employing the eigenvector properties becomes

$$\mu N \mathbf{u}_H^T \mathbf{u}^{(1)} + \mathbf{u}_H^T \mathbf{X} \mathbf{u}_C = \nu N \mathbf{u}_H^T \mathbf{u}^{(1)}. \quad (41)$$

Rearranging and noting that $\mathbf{u}_H^T \mathbf{u}^{(1)} = \mathbf{u}_H^T \mathbf{u}_{\parallel}^{(1)}$, we find

$$\mathbf{u}_H^T \mathbf{u}^{(1)} = \mathbf{u}_H^T \mathbf{u}_{\parallel}^{(1)} = \frac{\mathbf{u}_H^T \mathbf{X} \mathbf{u}_C}{\nu N - \mu N}, \quad (42)$$

which suggests the following solution for $\mathbf{u}_{\parallel}^{(1)}$,

$$\mathbf{u}_{\parallel}^{(1)} = \frac{[\mathbf{X} \mathbf{u}_C]_{\parallel}}{\nu N - \mu N}. \quad (43)$$

Writing $[\mathbf{X} \mathbf{u}_C]_{\parallel}$ as a decomposed projection onto $\mathbf{u}_C$ and $\mathbf{u}_H$,

$$[\mathbf{X} \mathbf{u}_C]_{\parallel} = (\mathbf{u}_C^T \mathbf{X} \mathbf{u}_C) \mathbf{u}_C + (\mathbf{u}_H^T \mathbf{X} \mathbf{u}_C) \mathbf{u}_H, \quad (44)$$

allows us to write $\mathbf{u}_{\parallel}^{(1)}$ as

$$\mathbf{u}_{\parallel}^{(1)} = \frac{(\mathbf{u}_C^T \mathbf{X} \mathbf{u}_C)}{\nu N - \mu N} \mathbf{u}_C + \frac{(\mathbf{u}_H^T \mathbf{X} \mathbf{u}_C)}{\nu N - \mu N} \mathbf{u}_H. \quad (45)$$

We now seek to solve for $\mathbf{u}_\perp^{(1)}$. Using the decomposition Eq. (39) in the first-order equation, Eq. (33), gives

$$\langle \mathbf{A} \rangle [\mathbf{u}_\parallel^{(1)} + \mathbf{u}_\perp^{(1)}] + \mathbf{X} \mathbf{u}_C = \nu N [\mathbf{u}_\parallel^{(1)} + \mathbf{u}_\perp^{(1)}] + \lambda^{(1)} \mathbf{u}_C. \quad (46)$$

Retaining only the terms that have components orthogonal to the $\mathbf{u}_C, \mathbf{u}_H$ plane and rearranging yields

$$\mathbf{u}_\perp^{(1)} = \frac{[\mathbf{X} \mathbf{u}_C]_\perp}{\nu N}. \quad (47)$$

Using the substitution $[\mathbf{X} \mathbf{u}_C]_\perp = \mathbf{X} \mathbf{u}_C - [\mathbf{X} \mathbf{u}_C]_\parallel$ then gives the solution for the orthogonal component

$$\mathbf{u}_\perp^{(1)} = \frac{\mathbf{X} \mathbf{u}_C - [(\mathbf{u}_C^T \mathbf{X} \mathbf{u}_C) \mathbf{u}_C + (\mathbf{u}_H^T \mathbf{X} \mathbf{u}_C) \mathbf{u}_H]}{\nu N}. \quad (48)$$

Combining Eqs. (45) and (48) gives the solution for the first-order perturbation to the eigenvector $\mathbf{u}^{(1)}$,

$$\mathbf{u}^{(1)} = \frac{\mathbf{X} \mathbf{u}_C}{\nu N} + \frac{\mu(\mathbf{u}_C^T \mathbf{X} \mathbf{u}_C)}{\nu N(\nu - \mu)} \mathbf{u}_C + \frac{\mu(\mathbf{u}_H^T \mathbf{X} \mathbf{u}_C)}{\nu N(\nu - \mu)} \mathbf{u}_H. \quad (49)$$

C. Second-order treatment

Having found the first-order eigenvalue and eigenvector perturbations, $\lambda^{(1)}$ and $\mathbf{u}^{(1)}$, we can now solve for the second-order correction $\lambda^{(2)}$. We multiply both sides of Eq. (34) by $\mathbf{u}_C^T$ and then solve to get

$$\lambda^{(2)} = \mathbf{u}_C^T \mathbf{X} \mathbf{u}^{(1)} - \lambda^{(1)} \mathbf{u}_C^T \mathbf{u}^{(1)} \quad (50)$$

$$= \frac{\mathbf{u}_C^T \mathbf{X}^2 \mathbf{u}_C}{\nu N} - \frac{(\mathbf{u}_C^T \mathbf{X} \mathbf{u}_C)^2}{\nu N} + \frac{\mu(\mathbf{u}_H^T \mathbf{X} \mathbf{u}_C)^2}{\nu N(\nu - \mu)}, \quad (51)$$

where Eqs. (36) and (49) have been used to obtain the second line.

We seek the expected value $\langle \lambda^{(2)} \rangle$ and consider the right-hand terms of Eq. (51) in succession. Expanding the expected value of the first term yields

$$\begin{aligned} \frac{\langle \mathbf{u}_C^T \mathbf{X}^2 \mathbf{u}_C \rangle}{\nu N} &= \frac{1}{\nu N} \frac{1}{N} \left\{ \sum_{i,j=1}^{\frac{N}{2}} \langle (\mathbf{X}^2)_{ij} \rangle + \sum_{i,j=\frac{N}{2}+1}^N \langle (\mathbf{X}^2)_{ij} \rangle \right. \\ &\quad \left. - 2 \sum_{i=1}^{\frac{N}{2}} \sum_{j=\frac{N}{2}+1}^N \langle (\mathbf{X}^2)_{ij} \rangle \right\}, \end{aligned} \quad (52)$$

where $(\mathbf{X}^2)_{ij} = \sum_{k=1}^N X_{ik} X_{kj}$. As the elements of $\mathbf{X}$ are independent, the cross-element terms in this sum have vanishing expectation: $\langle X_{ik} X_{kj} \rangle = 0$ for $i \neq j$. When $i = j$, the value of $\langle X_{ik}^2 \rangle$ is either the ingroup or outgroup variance: $\langle X_{ik}^2 \rangle = \sigma_{\text{in}}^2$ if $i, k \leq N/2$ or $i, k > N/2$; $\langle X_{ik}^2 \rangle = \sigma_{\text{out}}^2$ otherwise. Accordingly, the expectations for the elements of $\mathbf{X}^2$ are given by

$$\langle (\mathbf{X}^2)_{ij} \rangle = \begin{cases} \frac{N}{2} \sigma_{\text{in}}^2 + \frac{N}{2} \sigma_{\text{out}}^2 = N \sigma^2, & i = j, \\ 0, & i \neq j. \end{cases} \quad (53)$$

The above equation reduces the double sums in the first two terms in Eq. (52) to single sums over $\langle (\mathbf{X}^2)_{ii} \rangle = N \sigma^2$, which can then be combined. The last term, which contains only

off-diagonal elements of $\langle (\mathbf{X}^2) \rangle$, vanishes. The contribution of the first term in Eq. (51) to $\langle \lambda^{(2)} \rangle$ is therefore

$$\frac{\langle \mathbf{u}_C^T \mathbf{X}^2 \mathbf{u}_C \rangle}{\nu N} = \frac{1}{\nu N} \frac{1}{N} \sum_{i=1}^N N \sigma^2 \quad (54)$$

$$= \frac{\sigma^2}{\nu}. \quad (55)$$

We now turn to the second and third terms on the right-hand side of Eq. (51). The numerator of the second term, $(\mathbf{u}_C^T \mathbf{X} \mathbf{u}_C)^2$, involves the square of $\lambda^{(1)}$ by Eq. (36). Hence, its expected value is equivalent to the variance of $\lambda^{(1)}$ (which has zero mean), and so goes as σ^2 as shown above. Consequently, the expected value of the second term goes as σ^2/N . The same argument holds for the third term as its numerator depends on $\mathbf{u}_H^T \mathbf{X} \mathbf{u}_C$, which is likewise tantamount to the sum over the random ingroup and outgroup variables. Therefore, in the large- N regime of concern here, the second and third terms, which have a $1/N$ dependence, can be neglected in comparison with the first term, which is independent of N . Accordingly, Eq. (55) gives the second-order eigenvalue perturbation,

$$\langle \lambda^{(2)} \rangle = \frac{\sigma^2}{\nu}. \quad (56)$$

Having found that the first-order perturbation vanishes on average and given the second-order perturbation above, we arrive at the approximate solution for the expected value of the contrast eigenvalue of Eq. (28),

$$\langle \lambda_C \rangle = \nu N + \frac{\sigma^2}{\nu}. \quad (57)$$

Since the ratio of the second-order correction to the unperturbed eigenvalue goes as σ^2/N , the expansion parameter, given by its square root, is therefore $\mathcal{O}(\sigma/\sqrt{N})$ as stated at the beginning of this calculation.

A similar calculation, this time expanding about the homogeneous eigenvector $\mathbf{u}_H$, gives us the solution for λ_H , which simply involves swapping out ν for μ in the preceding equation,

$$\langle \lambda_H \rangle = \mu N + \frac{\sigma^2}{\mu}. \quad (58)$$

The analytical expressions for the signal eigenvalues, Eqs. (57) and (58), are plotted in Fig. 3(a) for values outside the main spectral band. They are observed to be in good agreement with the outlying eigenvalues of the numerical spectrum. This is the case even though, rather than an average over many generated networks, each horizontal slice represents just one instance, a reflection of the self-averaging behavior of large random networks. These expressions also work well in the nonsparse limit as shown in Fig. 5 for the case where the contrast eigenvalue becomes the leading eigenvalue beyond the assortative transition. The predicted λ_C is observed to separate from the spectral edge past a critical density, which decreases with network size, and shows good agreement with the first eigenvalue of the simulated network, particularly for the two larger networks. For the special case $\mu = \nu$, the same signal eigenvalue expressions still hold but we note how this case affects the derivation. When $\mu \neq \nu$, the last term on the right-hand side of Eq. (51) could be neglected above as

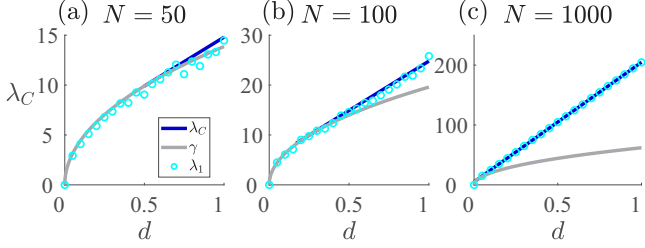


FIG. 5. Theoretical contrast [λ_C , Eq. (57)] and band edge [γ , Eq. (A15)] eigenvalues, along with the simulated leading eigenvalue (λ_1), as a function of network density. (a) $N = 50$, (b) $N = 100$, (c) $N = 1000$. λ_C is plotted for network density values where it meets or exceeds γ . One leading eigenvalue instance is computed for each d value. $d = d_{\text{in}} = d_{\text{out}}$, $p_{\text{in}}^+ = 0.6$, and $p_{\text{out}}^+ = 0.4$.

it is a factor of $1/N$ smaller than the first term. For $\mu = \nu$, however, the third term is singular. Yet, Eq. (41) implies that $\mathbf{u}_H^T \mathbf{X} \mathbf{u}_C = 0$ for $\mu = \nu$ (or is higher order for $\mu - \nu \sim 1/N$) and so the seemingly problematic third term does not arise. For the special cases $\nu = 0$ or $\mu = 0$, $\langle \lambda_C \rangle = 0$ or $\langle \lambda_H \rangle = 0$, respectively, and $\langle \mathbf{A} \rangle$ becomes rank 1. Figure 7(b) shows example spectra for $\mu = 0$ in which λ_C is the only signal eigenvalue.

The symmetric forms of the expressions for λ_C and λ_H reflect the fact that an orthogonal transformation $\mathbf{K}$ exists that transforms the contrast and homogeneous eigenvectors of $\langle \mathbf{A} \rangle$ into each other, that is, $\mathbf{u}_H = \mathbf{K} \mathbf{u}_C$ and $\mathbf{u}_C = \mathbf{K} \mathbf{u}_H$ where the primes denote the transformed system. Specifically, $\mathbf{K}$ is the diagonal matrix $\text{diag}(1, \dots, 1, -1, \dots, -1)$ where the negative values start at index $N/2 + 1$. It is its own inverse, $\mathbf{K}^{-1} = \mathbf{K}$. The transformation of the expected adjacency matrix, $\langle \mathbf{A}' \rangle = \mathbf{K} \langle \mathbf{A} \rangle \mathbf{K}$, flips the sign of the off-diagonal blocks so that $\langle A'_{\text{out}} \rangle = -\langle A_{\text{out}} \rangle$. Therefore, by Eqs. (11) and (12), $\mu' = \nu$ and $\nu' = \mu$, which swaps the perturbed signal eigenvalues, $\lambda'_H = \lambda_C$ and $\lambda'_C = \lambda_H$.

An alternative calculation of the signal eigenvalues based on random matrix theory and complex analysis is presented in the Appendix B. We find the same formulas for the signal eigenvalues [see Eq. (B29)] as have been derived here.

V. TRANSITION BOUNDARIES

In this section, we derive theoretical predictions for the boundaries of the detectability and sociality transitions. As discussed in Sec. III, these transitions occur when the signal eigenvalues merge with the main spectral band. From Eq. (A15), the edges of the main band of $\mathbf{X}$ are given by $\pm 2\sigma\sqrt{N}$, a formula that is a straightforward adaptation of the band edge of Wigner's semicircle distribution. We consider the community detectability transitions first, that is, those involving the contrast eigenvector $\mathbf{u}_C$, whose eigenvalue is given by Eq. (57). The detectability transition will therefore occur when

$$\nu N + \frac{\sigma^2}{\nu} = 2\sigma\sqrt{N}. \quad (59)$$

Solving for the critical value ν_* yields

$$\nu_* = \frac{\sigma(\nu_*)}{\sqrt{N}}. \quad (60)$$

The notation $\sigma(\nu_*)$ serves as a reminder that ν_* also appears on the righthand side due to the functional dependence of σ given by Eq. (24). The community structure is detectable when $|\nu| > \nu_*$. In particular, the assortative transition occurs for $\nu = \nu_*$ and the detectability transition for disassortative structure occurs for $\nu = -\nu_*$.

We observe that the transition condition Eq. (60) also results by setting the noise power equal to the signal power. Defining the noise power as the projection of $\mathbf{X}^2$ onto $\mathbf{u}_C$, its average, $\langle \mathbf{u}_C^T \mathbf{X}^2 \mathbf{u}_C \rangle$, is found using Eq. (55) to be $N\sigma^2$. One could also arrive at this value by considering how much of the total noise variance, $N^2\sigma^2$, is carried on average by each of N randomly chosen orthogonal basis vectors. Equating the signal power to the average noise power, $\nu^2 N^2 = N\sigma^2$, yields Eq. (60).

The sociality transitions associated with the homogeneous signal occur when the homogeneous eigenvalue given by Eq. (58) equals the band edge eigenvalue. This yields a critical value μ_* ,

$$\mu_* = \frac{\sigma(\mu_*)}{\sqrt{N}}. \quad (61)$$

The prosocial transition occurs for $\mu = \mu_*$ and the antisocial transition occurs for $\mu = -\mu_*$.

We now unpack the transition conditions derived above to express them in alternative ways in parameter space that will further intuitive understanding of the transition behavior and allow for connection with simulation results.

First, we substitute Eq. (24) for σ^2 in the detectability condition Eq. (60) to yield

$$N\nu^2 = \frac{1}{2}(d_{\text{in}} + d_{\text{out}}) - \mu^2 - \nu^2. \quad (62)$$

We point out that $(d_{\text{in}} + d_{\text{out}})/2$ is simply the overall tie density in the network. For a sparse network, $d_{\text{in}}, d_{\text{out}} \ll 1$, we can neglect the μ^2 and ν^2 terms on the right-hand side, so that the detectability transitions occur at

$$\pm \nu_* = \pm \sqrt{\frac{1}{2N}(d_{\text{in}} + d_{\text{out}})}. \quad (63)$$

The positive sign corresponds to the assortative transition and the negative sign corresponds to the disassortative transition. As ν can be regarded as a natural parameter for the community structure, this structure (assortative or disassortative) becomes easier to detect as the network becomes more sparse since ν_* shifts to smaller values (but care should be taken to distinguish the behavior of ν from that of the affinities and animosities, which can behave oppositely with density as in Eq. (72) below). Weaker structure is also more detectable as the size of the network grows, as was already apparent from Eq. (60).

We now substitute into Eq. (62) the definitions Eqs. (11) and (12) for μ and ν and rearrange to obtain

$$0 = \langle A_{\text{out}} \rangle^2 - \frac{2N}{N+2} \langle A_{\text{out}} \rangle \langle A_{\text{in}} \rangle + \langle A_{\text{in}} \rangle^2 - \frac{2}{N+2} (d_{\text{in}} + d_{\text{out}}), \quad (64)$$

which can be solved for the critical value of $\langle A_{\text{out}} \rangle$ (omitting the asterisk),

$$\langle A_{\text{out}} \rangle = \frac{N}{N+2} \langle A_{\text{in}} \rangle \pm \sqrt{\frac{2}{N+2} (d_{\text{in}} + d_{\text{out}}) - \frac{4(N+1)}{(N+2)^2} \langle A_{\text{in}} \rangle^2}. \quad (65)$$

To write the detectability transitions completely in terms of the block model probabilities, we substitute Eqs. (9) and (10) for $\langle A_{\text{in}} \rangle$ and $\langle A_{\text{out}} \rangle$ into Eq. (65). Solving for the outgroup animosity and taking the large N limit yields

$$p_{\text{out}}^- = \frac{1}{2} - \frac{d_{\text{in}}}{d_{\text{out}}} \left(p_{\text{in}}^+ - \frac{1}{2} \right) \pm \frac{1}{d_{\text{out}}} \sqrt{\frac{d_{\text{in}} + d_{\text{out}} - 8d_{\text{in}}^2 (p_{\text{in}}^+ - \frac{1}{2})^2}{2N}}, \quad (66)$$

where the positive sign corresponds to the assortative transition. Neglecting the second term inside the square root yields the sparse limit, equivalent to Eq. (63).

For the sociality transitions, the sparse limit results in the condition

$$\pm \mu_* = \pm \sqrt{\frac{1}{2N} (d_{\text{in}} + d_{\text{out}})}. \quad (67)$$

The positive and negative signs correspond to the prosocial and antisocial transitions, respectively. The critical value of the outgroup animosity is given by

$$p_{\text{out}}^- = \frac{1}{2} + \frac{d_{\text{in}}}{d_{\text{out}}} \left(p_{\text{in}}^+ - \frac{1}{2} \right) \pm \frac{1}{d_{\text{out}}} \sqrt{\frac{d_{\text{in}} + d_{\text{out}} - 8d_{\text{in}}^2 (p_{\text{in}}^+ - \frac{1}{2})^2}{2N}}, \quad (68)$$

where the negative sign is used for the prosocial transition.

VI. STRUCTURAL BALANCE DYNAMICS

In its simplest incarnation, structural balance theory considers the stability of triads. Triads with all positive ties (“the friend of my friend is my friend”) or two negative ties (“the enemy of my enemy is my friend”) are considered balanced and so stable. In contrast, a triad with an odd number of negative ties will be unbalanced. For fully connected networks, assuming that all triads must be balanced over time implies that the system achieves either a state of global harmony in which all nodes are positively connected or two hostile camps with positive connections within each camp and negative connections between them [26]. Empirical signed networks in social systems such as international relations, student relationships, and online social networks have been found to be approximately balanced [1,3,27], exhibiting a tendency toward partition into two factions.

Although the concept of balance can be extended to arbitrary-length cycles, the triadic notion has motivated the construction of dynamical systems models that evolve the relationship between a pair of nodes as a function of their relationships with their network neighbors [18,19,28]. As

noted when Eq. (1) was introduced, if both members of a dyad have a positive relationship with a third node, then that will act toward making the focal dyad’s relationship more positive. In contrast, having oppositely signed relationships with the third node will contribute a force pulling the dyad toward a more conflictual relationship. Unbalanced triads wither away under these dynamics. Building upon Ref. [18], Marvel *et al.* [19] demonstrated that the model of structural balance dynamics defined by Eq. (1) almost always achieves a balanced state starting from random initial conditions. Equation (1) can be written as a matrix equation,

$$\frac{d\mathbf{Y}}{dt} = \mathbf{Y}^2, \quad (69)$$

where $\mathbf{Y}$ is the matrix of signed and continuous connectivity values, Y_{ij} , between node pairs. In support of its empirical relevance, Ref. [19] found that when implemented upon the initial network of several real world systems, this model well predicts the observed final network.

Equation (69) is the matrix form of a Riccati equation and has the following closed form solution [19]:

$$\mathbf{Y}(t) = \mathbf{Y}(0)[\mathbf{I} - \mathbf{Y}(0)t]^{-1}. \quad (70)$$

The elements Y_{ij} diverge to positive or negative infinity in a finite time t_f and so the solution only holds for $t < t_f$.

For the purposes of analyzing community structure, we convert the connectivity matrix $\mathbf{Y}$ to an adjacency matrix $\mathbf{A}$ with discrete values ± 1 and 0 by taking the sign of the connectivity values so that $A_{ij} = \text{sgn}(Y_{ij})$. The leading eigenvector of the initial connectivity matrix, $\mathbf{Y}(0)$, grows fastest and so dominates the solution as $t \rightarrow t_f$. As a result, the final adjacency matrix $\mathbf{A}_f$ corresponding to $\mathbf{Y}$ as $t \rightarrow t_f$ converges to the outer product,

$$\mathbf{A}_f = \mathbf{u}_1 \mathbf{u}_1^T, \quad (71)$$

where $\mathbf{u}_1$ consists of the signs of the leading eigenvector of $\mathbf{Y}(0)$.

The rank 1 structure of $\mathbf{A}_f$ toward which the connectivity matrix converges implies that the final network must partition into either two hostile factions or one harmonious community as consistent with the expectations of structural balance theory [19]. The final network consists of a single harmonious faction if the components of $\mathbf{u}_1$ are of uniform sign, but consists of two hostile factions if $\mathbf{u}_1$ contains both positive and negative values [19]. Note that these results hold only if there is a single dominant eigenvalue and the graph is connected. When $d_{\text{out}} = 0$, the graph is disconnected and the isolated identity blocks will evolve independently of each other and the connectivity matrix will become rank 2 after structural balance dynamics. Although the first two eigenvalues of $\langle \mathbf{A} \rangle$ are equal in the $p_{\text{out}}^+ = 1/2$ case as well, the network remains connected and, due to stochasticity, one eigenvalue will inevitably be slightly larger in the realized $\mathbf{A}$, causing it to generate the ultimate rank 1 state. The case of $d_{\text{in}} = 0$ is also connected and so evolves to rank 1.

We will investigate the evolution of networks with community structure under the structural balance model (69). The initial connectivity matrix is taken to be proportional to an initial adjacency matrix $\mathbf{A}_0$ generated using the stochastic block model, in particular $\mathbf{Y}(0) = \mathbf{A}_0/N$. Figure 6(a) shows an

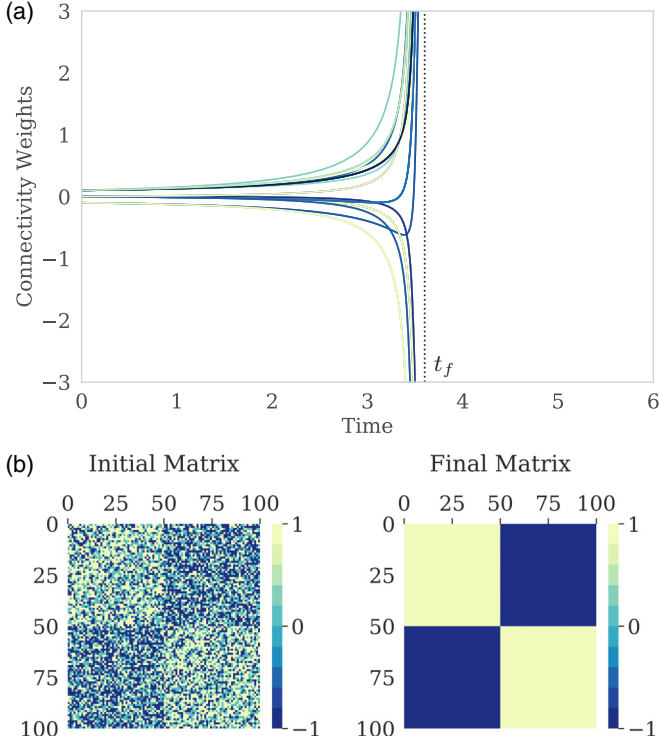


FIG. 6. Structural balance evolution of a network with community structure. (a) Connectivity network weights $Y_{ij}(t)$ over time evolved by Eq. (69) from an initial network generated by a stochastic block model (with values rescaled by $1/N$). (b) Initial and final adjacency matrices. Initial matrix parameters $N = 100$, $d_{in} = 0.7$, $d_{out} = 0.7$, $p_{in}^+ = 0.65$, and $p_{out}^+ = 0.35$.

example of the evolution of network connectivity values over time from a $\mathbf{Y}(0)$ corresponding to the $\mathbf{A}_0$ shown on the left in Fig. 6(b). We see that the $Y_{ij} \rightarrow \pm\infty$ and the final adjacency matrix $\mathbf{A}_f$ on the right shows the split into two factions with positive ties within each faction and negative ties between factions. It is given by the outer product Eq. (71) with $\mathbf{u}_1 = \mathbf{u}_C$ so that the factions correspond with the identity blocks A and B .

VII. STRUCTURAL BALANCE BEHAVIORAL REGIMES

As the final structure of these dynamical networks is dominated by the initial network's leading eigenvector, we can determine the extent to which networks in our parameter space will become assortative or homogeneous using our transition formulas derived above. First we treat a special case before exploring more general parameters.

A. Ingroup affinity equals outgroup animosity

We consider the simple case in which the ingroup affinity is set equal to the outgroup animosity, $p_{in}^+ = p_{out}^-$. We also make the simplification $d_{in} = d_{out} = d$. For this case, $\langle A_{in} \rangle = -\langle A_{out} \rangle$ so that $\mu = 0$, i.e., there is no homogeneous signal, and $\nu = -\langle A_{out} \rangle$. Using $\langle A_{out} \rangle = d(1 - 2p_{out}^-)$ in Eq. (63), we

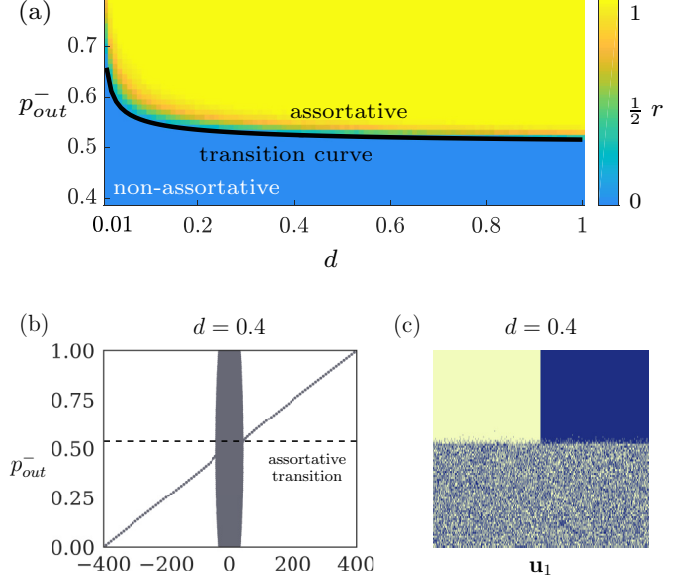


FIG. 7. (a) Assortativity of final adjacency matrix as a function of d and p_{out}^- for network size $N = 1000$ averaged over 400 simulations. Solid curve is the theoretical transition boundary given by Eq. (72). Note that the theoretical curve corresponds to the early part of the transition in which r just begins to rise, whereas the more visually distinctive yellow-cyan interface marks the middle of the transition. (b) Initial adjacency matrix spectra for $d = 0.4$ and increasing p_{out}^- . Dashed line indicates theoretical transition point. (c) Component signs for the leading eigenvector $\mathbf{u}_1$ of the initial adjacency matrix. The disassortative transition is not dynamically relevant so only the upper part of the p_{out}^- scale is plotted in (a).

solve for the critical outgroup animosity

$$p_{out}^- = \frac{1}{2} \left(1 + \sqrt{\frac{1}{dN}} \right). \quad (72)$$

Note that we only use the positive sign from Eq. (63), since it is the assortative transition, not the disassortative one, that involves the leading eigenvector.

Figure 7 shows the alignment between the dynamical regimes evolved by the structural balance model and the community structure of the initial network. Figure 7(a) plots the assortativity r defined by Eq. (6) of the final adjacency matrix, $\mathbf{A}_f$ as averaged over 400 initial networks generated by the stochastic block model at each point in the d and p_{out}^- parameter space. The yellow region represents the fully assortative outcome where the system evolves into two factions corresponding to the identities defined by the stochastic block model. In the blue region where $r \approx 0$, the two final factions are well mixed by identity. We see that the boundary between these two regions is in good accord with Eq. (72). As the network density increases, the assortative regime is observed to grow, extending down to smaller outgroup animosity values. Figures 7(b) and 7(c) plot, respectively, the spectrum and first eigenvector of the initial adjacency matrix, $\mathbf{A}_0$ for a constant density value. The transition in the structural balance dynamics mirrors the behavior of the first eigenvector which undergoes an assortative detectability transition at $p_{out}^- = 0.525$. These plots also confirm that the

leading eigenvector is never homogeneous and so one does not expect to observe the single faction outcome in this case (being extremely improbable).

B. General parameter conditions

We now analyze the final states of networks generated using more general parameter conditions. We plot the behavior of the signal transition curves in the two-dimensional parameter space defined by p_{in}^+ and p_{out}^- for fixed values of d_{in}/d_{out} .

As the homogeneous signal was irrelevant in the previous case, we needed only plot the assortativity r . However, the prosocial transition will occur in general and so we must measure the extent to which nodes can be found in one large group. We define the homogeneity h as the fraction of all nodes that can be assigned to a single group by virtue of having a common sign in the leading eigenvector of the adjacency matrix. When all nodes have a common sign, they are positively connected to all other nodes so that $h = 1$, while when nodes are divided into two equal factions, the homogeneity assumes its minimum value, $h = 1/2$.

The top plots of Fig. 8 show the assortativity and homogeneity of the final adjacency matrix evolved by the structural balance dynamics in the parameter space defined by the ingroup affinity and outgroup animosity. They can then be linearly combined to effect their joint visualization as shown in the bottom plot. The assortative transition boundary predicted by Eq. (66) separates the assortative from nonassortative two-faction states while the prosocial theoretical boundary of Eq. (68) separates homogeneous single-faction states from the nonassortative two-faction states. These regimes relate to the initial network spectrum as follows: the blue region is where the homogeneous eigenvalue is both the largest eigenvalue and outside the main band; the cyan region is where the leading eigenvalue is part of the main band; and the yellow is where the contrast eigenvalue is largest and outside the main band.

The horizontal yellow-blue interface observed in Fig. 8(c) for larger p_{in}^+ values corresponds to the signal crossing transition in which the homogeneous and contrast eigenvectors exchange places, which occurs outside the noise band (see Fig. 4). Equating the contrast and homogeneous eigenvalue expressions, Eqs. (57) and (58), we find that the transition occurs when $\nu = \mu$, which implies that $\langle A_{out} \rangle = 0$ or equivalently $p_{out}^- = 0.5$.

Figure 9 shows how the density ratio d_{in}/d_{out} and overall network density $d = (d_{in} + d_{out})/2$ affect the assortative and prosocial transition curves. As d_{in}/d_{out} increases, the transition curves become steeper, implying that denser regions of the connectivity network have more influence on the final structure than sparse regions. Figure 10 shows how network size affects the location and shape of the assortative and homogeneous transitions. As N increases, the transitions become sharper and more closely aligned with the theoretical prediction for the critical value of p_{out}^- .

VIII. DISCUSSION

In this section, we first make some observations concerning our results on community and, more broadly, network structure, a subject of relevance to both unsigned and signed

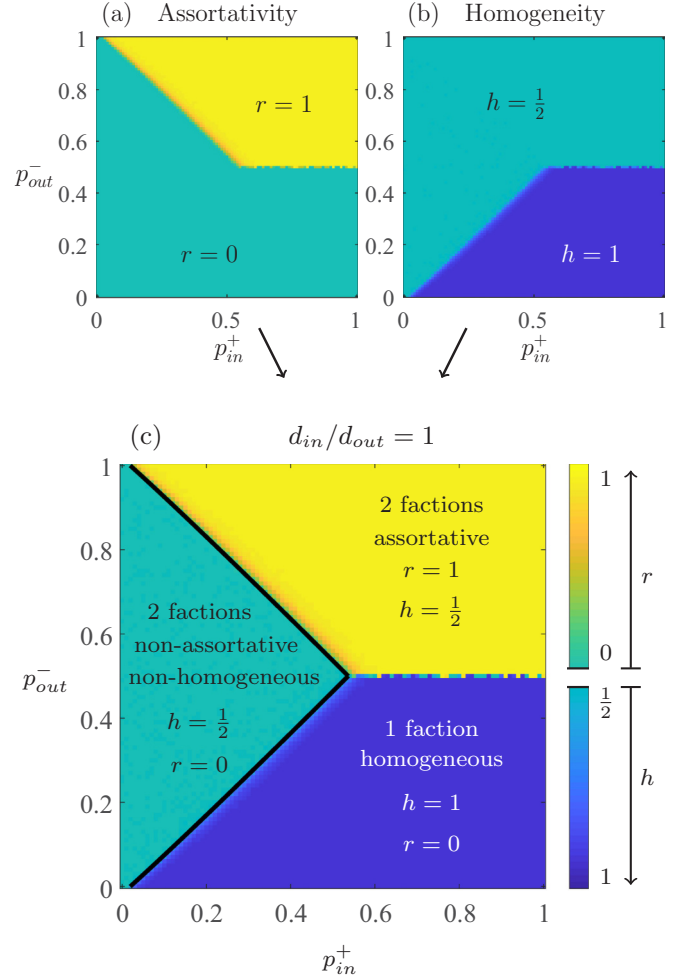


FIG. 8. (a) Assortativity and (b) homogeneity of final network states evolved by structural balance model Eq. (69) as a function of ingroup affinity and outgroup animosity. (c) Assortativity and homogeneity are mapped using the measure $z = r - 2h + 1$ to generate the joint heat map. For convenience, two separate color-bar scales are shown instead of z . The upper black curve indicates the assortative transition boundary, Eq. (66), while the lower black curve indicates the homogeneous transition boundary, Eq. (68). Heatmap values generated by averaging over four simulations for parameters $d_{in} = d_{out} = 0.45$ and $N = 1000$.

networks. We then turn to structural balance dynamics, an intrinsically signed network avenue of research. As it is particularly applicable to social systems, we speculate as to connections between our results and the dynamics of conflicts.

Our results can be applied to unsigned networks by taking the ingroup and outgroup animosities to be zero so that $p_{in}^+ = 1$ and $p_{out}^+ = 1$. In this case, our expressions for the homogeneous and contrast eigenvalues can be reduced to the sparse-limit forms reported in Refs. [13,14] for the two leading eigenvalues of the adjacency matrix. These can be obtained by neglecting the μ^2 and ν^2 contributions to σ^2 in Eq. (24) and then inserting into Eqs. (58) and (57) for the homogeneous and contrast eigenvalues, respectively. This yields the expressions (outside the noise band),

$$\lambda_H = \frac{N(d_{in} + d_{out})}{2} + 1, \quad (73)$$

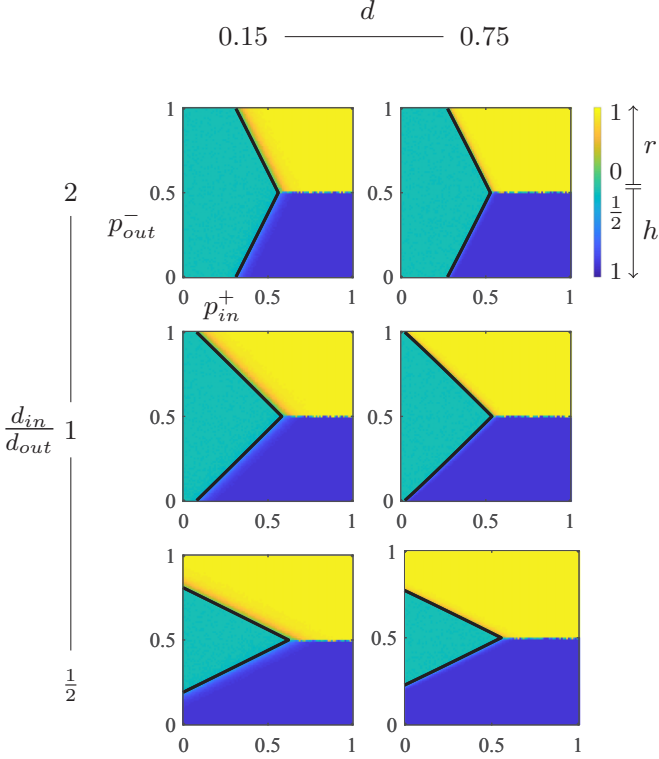


FIG. 9. Final state heat maps for increasing network density $d = (d_{\text{in}} + d_{\text{out}})/2$ and decreasing density ratio $d_{\text{in}}/d_{\text{out}}$. The assortativity and homogeneity are integrated via the z metric (see Fig. 8). Heatmap values generated by averaging over 4 simulations with $N = 1000$.

$$\lambda_C = \frac{N(d_{\text{in}} - d_{\text{out}})}{2} + \frac{d_{\text{in}} + d_{\text{out}}}{d_{\text{in}} - d_{\text{out}}}. \quad (74)$$

These forms, however, do not manifest the μ, ν interchange symmetry, an essential feature of the signed network case.

An important difference between unsigned and signed networks concerns which of the two leading eigenvectors outside the main spectral band may signify community structure. The above equations can be used to show that we can never observe the contrast eigenvalue as being larger than the homogeneous eigenvalue in the unsigned case. The sign of the difference, $\lambda_H - \lambda_C$, depends on the sign of $N - 2/(d_{\text{in}} - d_{\text{out}})$. The second term is equal to $1/\nu$, so that we must have $\nu < 1/N$ for the contrast eigenvalue to exceed the homogeneous eigenvalue. This condition upon ν in conjunction with Eq. (63) for the assortative transition in the sparse limit, which sets the minimum value of ν for λ_C to appear outside the noise band, then necessitates $\mu = (d_{\text{in}} + d_{\text{out}})/2 < 1/N$. However, this regime is below the threshold, $\mu_* = 1/N$, for the prosocial transition in the sparse limit as obtained from Eq. (67). Further, the righthand sides of Eqs. (63) and (67) are the same so that the assortative and prosocial transitions occur at the same critical value, $\nu_* = \mu_*$. For unsigned networks, $\nu \leq \mu$ and so if $\mu < \mu_*$, then $\nu < \nu_*$. Therefore, if λ_H is within the noise band then so must λ_C and hence we can never observe $\lambda_C > \lambda_H$.

Consequently, in unsigned networks, assortative community structure is represented by the second eigenvalue of the

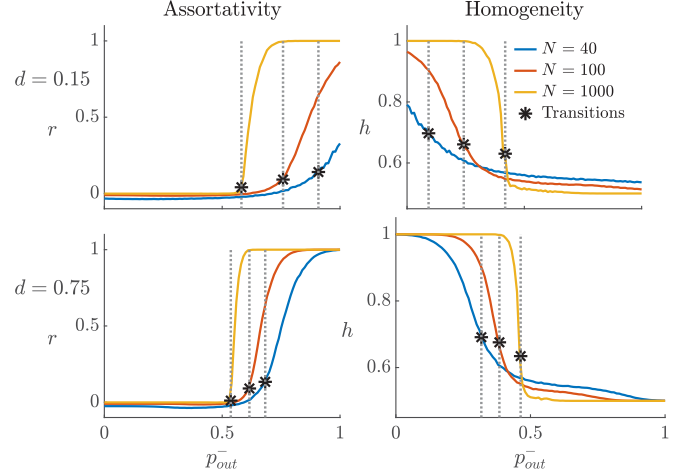


FIG. 10. Assortativity and homogeneity as a function of p_{out}^- for networks of increasing size. Black stars mark the predicted values of p_{out}^- at the transition points as given by Eqs. (66) and (68) for the assortative and prosocial transitions, respectively. Parameter values are $p_{\text{in}}^+ = 0.5$, $d = 0.15, 0.75$, and $N = 40, 100, 1000$ averaged over 2000, 1000, and 10 trials, respectively.

adjacency matrix (when past the detectability threshold) but not the first. In signed networks, the ordering of μ and ν is not restricted and so the first eigenvalue may signify the community structure while the second, if above the prosocial transition, signifies the homogeneous structure. Relatedly, while the number of outlying eigenvalues is equal to the number of communities in unsigned networks [8], this need not be the case in signed networks. For instance, the case of equal ingroup affinity and outgroup animosity treated in Sec. VII has only one outlying eigenvalue but two communities.

The sociality transitions, which involve the homogeneous eigenvector, bear upon the question of whether a network exhibits a propensity toward positive versus negative tie formation, a question that is unique to signed networks. In unsigned networks, the prosocial transition is present but its significance corresponds to the emergence of a giant connected component; Eq. (67) yields a transition condition of $N(d_{\text{in}} + d_{\text{out}})/2 = 1$ in accordance with the result for a giant component with two-community structure noted in Sec. II A. Typically, community structure is taken to connote the existence of multiple communities as it is linked to the community detection problem and the assignment of nodes to communities. The sociality transitions, which are not relevant to the community assignment problem, involve network structure more generally rather than community structure per se. However, the sociality transitions do lend themselves to a sense by which a signed network can be viewed as forming a single community: the existence of a global tendency, irrespective of ingroup and outgroup distinctions, toward the formation of positive versus negative ties. The prosocial transition provides a spectral signature for concluding that the network forms a single community in which relationships are generally friendly or cooperative. Conversely, the antisocial transition provides a signature of a single (anti)community marked by hostility, a Hobbesian

state of all against all. The intermediate case, where the homogeneous eigenvalue does not appear outside the noise band, can be taken as marking the absence of a single community with a definitive inclination toward positive or negative tie formation.

Turning to the dynamics, taking networks generated by the stochastic block model and evolving them under the structural balance model entails a dramatic shift in the processes governing network evolution (beyond the change from stochastic to deterministic): a process in which tie formation in a dyad depends only upon its ingroup or outgroup identity is replaced by one where the friendly or hostile relationships among mutual neighbors drives tie value evolution regardless of identity. One rationale for such a shift can be provided by assuming that there is a qualitative change in the nature of the interactions. For instance, hostile relationships characterized by insulting words or gestures may be replaced by physical violence. A second rationale could involve, not a change in the nature of interactions, but a growing awareness that hostile interactions have the potential to become much more prevalent. For example, in a country or region containing two broad identity types, such as ethnicity or ideology, in which individuals or small gangs sporadically clash (either within or across identity lines), a sudden collapse of the central government may lead to a growing sense of looming systemic violence. In either of these rationales, nodes are motivated to seek and maintain allies so that another node's status as the enemy of an enemy or friend of a friend becomes a crucial determinant in relationship formation and evolution.

Our results may inform debates about the interplay of identity and power in conflicts under anarchy consisting of many actors such as insurgencies, civil wars, and international relations. For ethnic conflict, the shift in models discussed above is supported by the observation that the turn from nonviolent to violent conflict represents a qualitative change in dynamics [29]. In the literature on civil wars and ethnic violence, some theories stress mechanisms in which ethnic or religious identity plays an intrinsic role in producing high levels of polarization and violence along identity lines while other theories stress the role of microprocesses of conflict among local actors rather than a pre-existing identity schism [29–31]. As node identity plays no role in the network evolution, structural balance dynamics is consistent with the latter view. However, the sharp transition to the assortative state shows that the dynamics can lock in initial differences in identity-driven community structure even when they are not large, a behavior consistent with the observation that the polarization and violence in ethnic civil wars often appears to be disproportionate to the initial level of ethnic tension. But the existence of the nonassortative regime implies that identity polarization will not arise when the initial structure is sufficiently weak. Thus, conflict takes on an essentially binary nature in that it is either completely polarized by identity or not at all. Additionally, it has been argued that, contrary to some theories, there is no inherent difference in the dynamics between ideological and ethnic civil wars in terms of their potential for polarization and violence [30]. Our results are consistent with such a claim as it is the initial relationships that matter regardless of whether they are due to similar ethnicity or similar ideology.

IX. CONCLUSION

This paper has contributed to two distinct areas of signed network research—community structure and structural balance theory, linking them via the impact of the former upon the latter. We have elucidated the spectrum of unweighted and undirected signed networks generated by a two-community stochastic block model via two independent methods, perturbation analysis and, in the Appendices, a random matrix theory treatment that extends prior work on unsigned networks [13,14]. The expected matrix, $\langle \mathbf{A} \rangle$, in the block model can be decomposed into two signals—a homogeneous eigenvector, $\mathbf{u}_H$, related to the expected tie value, μ , over the network and a contrast eigenvector, $\mathbf{u}_C$, related to the half-difference, ν , between the expected ingroup and outgroup tie values and which encodes the community structure. These signal eigenvectors exhibit transitions at the points where they merge with the main spectral band associated with the noise produced by the zero-mean random matrix $\mathbf{X}$. There are four potential transitions corresponding to the intersections of the two signals with the positive and negative edges of the main band. For the contrast eigenvector, these intersections induce the assortative and disassortative transitions, respectively, and mark changes in community detectability. The homogeneous eigenvector undergoes sociality transitions, prosocial and antisocial, in which emergence from the noise band signifies an overall tendency toward the formation of cooperative or conflictual relationships, respectively, with other nodes.

We derived analytical expressions for the signal eigenvalues in the presence of the noise by performing a perturbation expansion in which the contributions from the noise $\mathbf{X}$ were treated as small corrections to the eigenvalues of $\langle \mathbf{A} \rangle$. Equations (26) and (27) reveal a second-order correction proportional to the average tie variance and are symmetric under the interchange of μ and ν . The same expressions are derived in the Appendices using random matrix theory along with the formula for the main band edges that is a straightforward modification of Wigner's semicircle law. The transition conditions, Eqs. (60) and (61), were determined by equating the signal eigenvalues to the band edge eigenvalues.

We investigated structural balance dynamics in the presence of initial community structure generated by the two-identity stochastic block model. These dynamics completely connect all nodes and allow for three broad regimes of final states: an assortative regime in which two hostile factions emerge that completely align with the two identity blocks; a nonassortative regime in which the two final factions are randomly composed with respect to identity; and a homogeneous regime consisting of a single harmonious faction with only positive ties. Since the dominant eigenvector of the initial network drives its structural balance dynamics and determines its final state, our spectral analysis allows us to chart the parameter conditions under which each of these states will emerge. The dynamical ascendance of the leading eigenvector implies that the regime boundaries occur where any two of the homogeneous signal, the contrast signal, and the noise band edge exchange places as the first eigenvalue. The assortative transition marks the boundary between the nonassortative and assortative regimes and the prosocial transition divides the nonassortative and homogeneous regimes. The boundary

between the homogeneous and assortative regimes represents a reversal in the ranking of the homogeneous and contrast eigenvalues rather than a transition involving the noise band. The theoretically predicted boundaries were found to agree with the simulation results obtained by solving the structural balance model over many random initial networks.

Finally, we note a few potential directions for future research. As with the unsigned case, spectral analysis of signed networks with community structure could be extended to systems with multiple communities, directed ties, and more realistic network statistics such as nonuniform degree distributions. The structural balance model we used is very simple and, problematically, leads to tie strengths which blow up in finite time. Accordingly, the extent to which the dynamical transitions we have identified persist for more realistic implementations of structural balance dynamics should be explored. More empirical work is also needed to understand the conditions under which real networks can be reasonably modeled by structural balance dynamics or variants thereof.

Code reproducing select results from this paper is available online [32].

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their helpful feedback. Support for this research was provided by the Big Data for Genomics and Neuroscience Training Grant under the National Institute of Health Grant No. 5T32LM012419-04, the Office of Naval Research under Grant No. N00014-16-1-2919, and by the Army Research Office under Grant No. W911NF1910291.

APPENDIX A: SPECTRUM OF RANDOM MATRIX $\mathbf{X}$

The signal eigenvalues of $\mathbf{A}$ are only visible if they are distinguishable from the noise in the system. In this section we find $\mathcal{O}(\|\mathbf{X}\|_2)$ as well as γ the spectral edge of the noise matrix $\mathbf{X}$ using random matrix theory.

We start by characterizing the distribution of eigenvalues. The empirical spectral distribution (e.s.d.) of a random matrix $\mathbf{X}$ is defined by

$$\rho(z) = \frac{1}{N} \sum_{i=1}^n \delta(z - \omega_i), \quad (\text{A1})$$

where ω_i are the eigenvalues of $\mathbf{X}$.

Wigner's semicircle distribution Eq. (A2) defines the eigenvalue density function for a symmetric random matrix size N with i.i.d. entries having variance m^2 ,

$$\phi(z) = \frac{1}{2\pi Nm^2} \sqrt{4Nm^2 - z^2}. \quad (\text{A2})$$

According to the Wigner limit theorem, the spectra of certain symmetric random matrices converge in distribution to Wigner's semicircle distribution [33,34], so that

$$\lim_{N \rightarrow \infty} \int_{-\infty}^c \rho(x) dx = \int_{-\infty}^c \phi(x) dx, \quad (\text{A3})$$

where c is any real number. Although the semicircle law originally applies only to random symmetric matrices with equal variances for all entries

[33,35,36], further inquiry has determined that random symmetric block Toeplitz matrices [37,38] also weakly converge to the semicircle law under certain conditions.

$\mathbf{X}$ is a random symmetric block Toeplitz matrix with distribution variances σ_{in}^2 and σ_{out}^2 in the on and off diagonal blocks, respectively, and therefore has a Wigner semicircle distribution of eigenvalues for some variance parameter m^2 which we have yet to determine. The edges of the semicircle enclose the band of eigenvalues in the interval $(-2m\sqrt{N}, 2m\sqrt{N})$. As $N \rightarrow \infty$, $\lambda_1(\mathbf{X}) = 2m\sqrt{N}$.

We find the variance parameter by bounding $\|\mathbf{X}\|_2$ with the Frobenius norm [39],

$$\frac{1}{\sqrt{N}} \|\mathbf{X}\|_F \leq \|\mathbf{X}\|_2 \leq \|\mathbf{X}\|_F, \quad (\text{A4})$$

where

$$\|\mathbf{X}\|_F = \sqrt{\sum_{j=1}^{\frac{N^2}{2}} |X_{\text{in}}|^2 + \sum_{j=1}^{\frac{N^2}{2}} |X_{\text{out}}|^2}. \quad (\text{A5})$$

The ingroup and outgroup sums are distributed as

$$\sum_{j=1}^{\frac{N^2}{2}} |X_{\text{in}}|^2 \sim \mathcal{N}\left(\frac{N^2}{2} \sigma_{\text{in}}^2, \frac{N^2}{2} \text{Var}(|X_{\text{in}}|^2)\right), \quad (\text{A6})$$

$$\sum_{j=1}^{\frac{N^2}{2}} |X_{\text{out}}|^2 \sim \mathcal{N}\left(\frac{N^2}{2} \sigma_{\text{out}}^2, \frac{N^2}{2} \text{Var}(|X_{\text{out}}|^2)\right). \quad (\text{A7})$$

The random variable Z_2 denoting the inside of the square root in Eq. (A5) is therefore distributed as

$$Z_2 = \sum_{j=1}^{\frac{N^2}{2}} |X_{\text{in}}|^2 + \sum_{j=1}^{\frac{N^2}{2}} |X_{\text{out}}|^2 \sim \mathcal{N}(N^2 \sigma^2, N^2 \xi^2), \quad (\text{A8})$$

$$\text{where } \xi^2 = \frac{\text{Var}(|X_{\text{in}}|^2) + \text{Var}(|X_{\text{out}}|^2)}{2}. \quad (\text{A9})$$

The expected value of Z_2 scales with $\sigma^2 N^2$ while the standard deviation only scales with ξN meaning $Z_2 = \mathcal{O}(\sigma^2 N^2)$. Therefore,

$$\|\mathbf{X}\|_F = \mathcal{O}(\sigma N). \quad (\text{A10})$$

Equation (A4) then implies that

$$\sigma \sqrt{N} \leq \|\mathbf{X}\|_2 \leq \sigma N. \quad (\text{A11})$$

Therefore, $\|\mathbf{X}\|_2$ must scale with σ as

$$\mathcal{O}(\|\mathbf{X}\|_2) \sim \sigma. \quad (\text{A12})$$

This result combined with the Wigner's semicircle distribution scaling implies that

$$\mathcal{O}(\|\mathbf{X}\|_2) = \sigma \sqrt{N}. \quad (\text{A13})$$

We continue the argument to find the leading eigenvalue of $\mathbf{X}$. We have found that our variance parameter must scale with σ , $m = \mathcal{O}(\sigma)$,

$$m = a\sigma = a\sqrt{\frac{\sigma_{\text{in}}^2 + \sigma_{\text{out}}^2}{2}}, \quad (\text{A14})$$

for some constant a .

We set $\sigma_{\text{in}}^2 = \sigma_{\text{out}}^2$ which implies that we have a traditional Wigner matrix with i.i.d. entries of variance σ_{in}^2 . This means $m = \sigma_{\text{in}} = \sigma$ and therefore $a = 1$. Now let $\sigma_{\text{in}}^2 \neq \sigma_{\text{out}}^2$ while keeping σ^2 constant. m scales only with σ and therefore must be still equal to σ . This means that for all variance values, $m = \sigma$.

This implies that the eigenvalue density function of $\mathbf{X}$ with a diagonal block variance of σ_{in}^2 and an off diagonal block variance of σ_{out}^2 is equivalent to the eigenvalue density function of a random matrix with uniform variance σ^2 . This substitution is made in previous derivations of the spectral band of unsigned stochastic block model matrices [13].

Thus, we have determined that the spectra of $\mathbf{X}$ has a Wigner's semicircle distribution with variance parameter σ . Therefore, the edge of the spectral band γ of $\mathbf{X}$ is

$$\gamma = 2\sigma\sqrt{N}. \quad (\text{A15})$$

APPENDIX B: SPECTRA OF A DERIVED FROM RANDOM MATRIX THEORY AND COMPLEX ANALYSIS

We will now use an alternative method to derive the spectra of $\mathbf{A}$ using random matrix theory and complex analysis. In this argument, we use the eigenvalues of the noise matrix $\mathbf{X}$, whose spectra we have defined in Appendix A, to find the eigenvalues of $\mathbf{X} + \nu N \mathbf{u}_C \mathbf{u}_C^T$. We then take these intermediate eigenvalues and use them to find the eigenvalues of $\mathbf{A} = \mathbf{X} + \nu N \mathbf{u}_C \mathbf{u}_C^T + \mu N \mathbf{u}_H \mathbf{u}_H^T$.

The expected adjacency matrix $\langle \mathbf{A} \rangle$ is given by Eq. (13). We consider the spectrum obtained by adding just the contribution of the contrast eigenvector, $\nu N \mathbf{u}_C \mathbf{u}_C^T$, to the noise matrix which yields the eigenvalue equation

$$(\mathbf{X} + \nu N \mathbf{u}_C \mathbf{u}_C^T) \mathbf{v} = z \mathbf{v}. \quad (\text{B1})$$

We wish to solve for the eigenvalues z , and so use the methods of Ref. [14] to convert Eq. (B1) into a trace representation Eq. (B6). We begin by rearranging the terms in Eq. (B1) to eliminate the eigenvector $\mathbf{v}$,

$$\mathbf{u}_C^T (z - \mathbf{X})^{-1} \mathbf{u}_C = \frac{1}{\nu N}. \quad (\text{B2})$$

The left-hand side of Eq. (B2) can be written as a sum by performing an eigenvector decomposition on $\mathbf{X}$,

$$\mathbf{u}_C^T \mathbf{S} (z \mathbf{I} - \mathbf{\Lambda})^{-1} \mathbf{S}^T \mathbf{u}_C = \frac{1}{\nu N}, \quad (\text{B3})$$

$$\sum_{i=1}^N \frac{(\mathbf{x}_i^T \mathbf{u}_C)^2}{z - \omega_i} = \frac{1}{\nu N}, \quad (\text{B4})$$

where $\mathbf{S} \mathbf{A} \mathbf{S}^T$ is the eigenvector decomposition of $\mathbf{X}$ and $\mathbf{x}_i$ are the eigenvectors of $\mathbf{X}$ (as well as the columns of $\mathbf{S}$). We find that $N(\mathbf{x}_i^T \mathbf{u}_C)^2 \sim \chi_1^2$ (the chi-square distribution) and therefore $\mathbb{E}[(\mathbf{x}_i^T \mathbf{u}_C)^2] = \frac{1}{N}$ and $\text{Var}[(\mathbf{x}_i^T \mathbf{u}_C)^2] = \frac{2}{N^2}$. This allows us to make the approximation $(\mathbf{x}_i^T \mathbf{u}_C)^2 = 1/N$ in Eq. (B4), giving us

$$\sum_{i=1}^N \frac{1/N}{z - \omega_i} = \frac{1}{\nu N}, \quad (\text{B5})$$

$$\frac{1}{N} \text{Tr}(z - \mathbf{X})^{-1} = \frac{1}{\nu N}. \quad (\text{B6})$$

We define $f(z)$ as follows:

$$f(z) = \text{Tr}(z - \mathbf{X})^{-1}. \quad (\text{B7})$$

The values of z that satisfy $f(z) = \frac{1}{\nu}$ are the eigenvalues of the matrix $\mathbf{X} + \nu N \mathbf{u}_C \mathbf{u}_C^T$. $f(z)$ has simple poles where $z = \omega_i$, $f(z) \rightarrow -\infty$ as $z \nearrow \omega_i$ and $f(z) \rightarrow \infty$ as $z \searrow \omega_i$. $f(z)$ is a continuous function within the interval $z \in [\omega_i, \omega_{i+1}]$, therefore for each interval $f(z) = \frac{1}{\nu}$ for some value $z \in [\omega_i, \omega_{i+1}]$. This means the eigenvalues z_i and ω_i are interlaced with the leading eigenvalue $z_1 > \omega_1$. The largest solution to $f(z) = \frac{1}{\nu}$ Eq. (B7) is the leading eigenvalue of $\mathbf{X} + \mu N \mathbf{u}_C \mathbf{u}_C^T$.

We now can repeat this process to find a formula for both leading eigenvalues by adding the homogeneous signal in addition to the contrast signal back into the noise matrix and solving for the resulting eigenvalues λ [14].

The new eigenvalue equation becomes

$$(\mathbf{X} + \nu N \mathbf{u}_C \mathbf{u}_C^T + \mu N \mathbf{u}_H \mathbf{u}_H^T) \mathbf{v} = \lambda \mathbf{v}. \quad (\text{B8})$$

We solve this equation for the eigenvalues λ_i using the same method used to solve Eq. (B1) and which is detailed in Ref. [14]. The resulting equation $g(\lambda)$ has a similar form to $f(z)$ but with an additional term,

$$\frac{1/N}{\lambda - z_1} + \sum_{i=2}^N \frac{1/N}{\lambda - z_i} = \frac{1}{\mu N}, \quad (\text{B9})$$

$$g(\lambda) = \frac{1}{\lambda - z_1} + \sum_{i=2}^N \frac{1}{\lambda - z_i}. \quad (\text{B10})$$

The values of λ that satisfy $g(\lambda) = \frac{1}{\mu}$ are the eigenvalues of the matrix $\mathbf{X} + \nu N \mathbf{u}_C \mathbf{u}_C^T + \mu N \mathbf{u}_H \mathbf{u}_H^T$. Without loss of generality, assume $\nu > \mu$, meaning we have added the largest eigenvalue mode to the noise matrix followed by the second largest eigenvalue mode. When $|\lambda - z_1| \gg \frac{1}{N}$, $\frac{1}{\lambda - z_1} = \mathcal{O}(1)$ and $\sum_{i=2}^N \frac{1}{\lambda - z_i} = \mathcal{O}(N)$. Because $\frac{1}{\lambda - z_1}$ is the dominant term only when $|\lambda - z_1| = \mathcal{O}(1/N^2)$ and the spectral values z_i that constitute the spectral band of $\mathbf{X} + \nu N \mathbf{u}_C \mathbf{u}_C^T$ are interlaced with the spectrum of $\mathbf{X}$, we may approximate $g(\lambda)$ with our previous function $f(\lambda) = \text{Tr}(\lambda - \mathbf{X})^{-1}$ for all λ values away from z_1 . $g(\lambda)$ has a singularity at z_1 meaning there is an additional solution to $g(\lambda) = \frac{1}{\mu}$ when $\lambda \approx z_1$.

Therefore, the signal eigenvalues, λ_H and λ_C are the largest magnitude solutions to $f(\lambda) = \frac{1}{\nu}$ and $f(\lambda) = \frac{1}{\mu}$. We can find an analytical form for $f(\lambda)$ by taking advantage of the Stieltjes transform representation of $\text{Tr}(\mathbf{X} - \lambda)^{-1}$.

The Stieltjes transform $S_\rho(\lambda)$ of density $\rho(t)$ is a function of the complex variable λ and is defined outside the real interval I ,

$$S_\rho(\lambda) = \int_I \frac{\rho(t)}{\lambda - t} dt, \quad \lambda \in \mathbb{C} \setminus I. \quad (\text{B11})$$

The normalized trace of $(\mathbf{X} - \lambda)^{-1}$ is equivalent to the Stieltjes transform of the spectral density of $\mathbf{X}$ [34],

$$\frac{1}{N} \text{Tr}(\mathbf{X} - \lambda)^{-1} = S_\rho(\lambda) = \int_I \frac{\rho(x)}{x - \lambda} dx, \quad (\text{B12})$$

where the e.s.d. has been previously defined, Eq. (A1). We may substitute the eigenvalue density function $\phi(x)$ for the

e.s.d. integrand in the Stieltjes transform since these functions converge in distribution [34],

$$S_\rho(\lambda) = \int_I \frac{\rho(x)}{x - \lambda} dx = \int_I \frac{\phi(x)}{x - \lambda} dx. \quad (\text{B13})$$

Substituting Eq. (A2) for $\phi(x)$ yields

$$S_\phi(\lambda) = \int_I \frac{\phi(x)}{x - \lambda} dx = \frac{1}{2\pi N\sigma^2} \int_{-2\sqrt{N}\sigma}^{2\sqrt{N}\sigma} \frac{\sqrt{4N\sigma^2 - x^2}}{x - \lambda} dx. \quad (\text{B14})$$

We solve this integral using multiple changes of variables. Our argument is adapted from previous work [34]. Letting $x = 2\sqrt{N}\sigma \cos(y)$, the above becomes

$$S_\phi(\lambda) = \frac{1}{\pi} \int_0^{2\pi} \frac{\sin^2(y)}{2\sqrt{N}\sigma \cos(y) - \lambda} dy \quad (\text{B15})$$

$$= \frac{1}{\pi} \int_0^{2\pi} \frac{\left(\frac{e^{iy} - e^{-iy}}{2i}\right)^2}{2\sqrt{N}\sigma \left(\frac{e^{iy} + e^{-iy}}{2}\right) - \lambda} dy. \quad (\text{B16})$$

The second change of variables is $\zeta = e^{iy}$, which gives

$$S_\phi(\lambda) = \frac{i}{4\pi\sigma\sqrt{N}} \oint_{|\zeta|=1} \frac{(\zeta^2 - 1)^2}{\zeta^2 \left(\zeta^2 - \frac{\lambda}{\sigma\sqrt{N}}\zeta + 1\right)} d\zeta, \quad (\text{B17})$$

$$\text{Let } h(\zeta) = \frac{(\zeta^2 - 1)^2}{\zeta^2 \left(\zeta^2 - \frac{\lambda}{\sigma\sqrt{N}}\zeta + 1\right)}. \quad (\text{B18})$$

The function $h(\zeta)$ has three poles,

$$\zeta_0 = 0, \quad (\text{B19})$$

$$\zeta_1 = \frac{\lambda + \sqrt{\lambda^2 - 4N\sigma^2}}{2\sigma\sqrt{N}}, \quad (\text{B20})$$

$$\zeta_2 = \frac{\lambda - \sqrt{\lambda^2 - 4N\sigma^2}}{2\sigma\sqrt{N}}. \quad (\text{B21})$$

We must determine which poles are inside the radius $|\zeta| = 1$ and then use the residue theorem to compute the integral. ζ_0 is a pole of order 2 and is inside the contour. Note that $\zeta_1\zeta_2 = 1$, and therefore if $|\zeta_2| \neq |\zeta_1|$ then only one of these poles can be inside the contour. We find, with the argument to follow, that ζ_2 is the pole inside the contour for λ values for which $\text{Im}(\lambda) > 0$ and ζ_1 is the pole inside the contour for λ values for which $\text{Im}(\lambda) < 0$.

We find that ζ_2 is the enclosed pole for $\text{Im}(\lambda) > 0$ by first supposing that $\text{Re}(\lambda) > 0$. It follows that $\text{Re}(\sqrt{\lambda^2 - 4N\sigma^2}) > 0$ and $\text{Im}(\sqrt{\lambda^2 - 4N\sigma^2}) > 0$, and therefore $|\lambda - \sqrt{\lambda^2 - 4N\sigma^2}| < |\lambda + \sqrt{\lambda^2 - 4N\sigma^2}|$, which reveals that $|\zeta_2| < |\zeta_1|$. Let us now suppose that $\text{Re}(\lambda) < 0$. It follows that $\text{Re}(\sqrt{\lambda^2 - 4N\sigma^2}) < 0$ and $\text{Im}(\sqrt{\lambda^2 - 4N\sigma^2}) > 0$, and therefore, by the same argument as above, once again $|\zeta_2| < |\zeta_1|$, meaning that ζ_2 is the enclosed pole.

We use a repetitive argument to show that ζ_1 is the enclosed pole for $\text{Im}(\lambda) < 0$. Supposing that $\text{Re}(\lambda) < 0$, we find that $\text{Re}(\sqrt{\lambda^2 - 4N\sigma^2}) > 0$ and $\text{Im}(\sqrt{\lambda^2 - 4N\sigma^2}) > 0$ which informs us that $|\zeta_1| < |\zeta_2|$. Now supposing that $\text{Re}(\lambda) > 0$, we find that $\text{Re}(\sqrt{\lambda^2 - 4N\sigma^2}) < 0$ and $\text{Im}(\sqrt{\lambda^2 - 4N\sigma^2}) > 0$ and therefore $|\zeta_1| < |\zeta_2|$, meaning that ζ_1 is the enclosed pole.

We now know that ζ_1 is inside the contour and ζ_2 is outside for $\text{Im}(\lambda) < 0$ while ζ_2 is inside the contour and ζ_1

is outside for $\text{Im}(\lambda) > 0$. We use the Residue theorem and the three poles Eqs. (B19)–(B21) to finish solving Eq. (B17) by integrating

$$\oint_{|\zeta|=1} h(\zeta) d\zeta = 2\pi i [\text{Res}(h, \zeta_0) + \text{Res}(h, \zeta_*)], \quad (\text{B22})$$

$$\text{where } \zeta_* = \begin{cases} \zeta_1, & \text{for } \text{Im}(\lambda) < 0, \\ \zeta_2, & \text{for } \text{Im}(\lambda) > 0. \end{cases} \quad (\text{B23})$$

Evaluating the residues yields

$$\oint_{|\zeta|=1} h(\zeta) d\zeta = 2\pi i \left[\frac{\lambda}{\sigma\sqrt{N}} \pm \frac{\sqrt{\lambda^2 - 4N\sigma^2}}{\sigma\sqrt{N}} \right]. \quad (\text{B24})$$

The $\pm$ in the above expression results from the two different residues for $\text{Im}(\lambda) > 0$ and $\text{Im}(\lambda) < 0$. We now multiply our result by the constant term to finish solving Eq. (B17),

$$S_\phi(\lambda) = \frac{-\lambda \pm \sqrt{\lambda^2 - 4N\sigma^2}}{2N\sigma^2}. \quad (\text{B25})$$

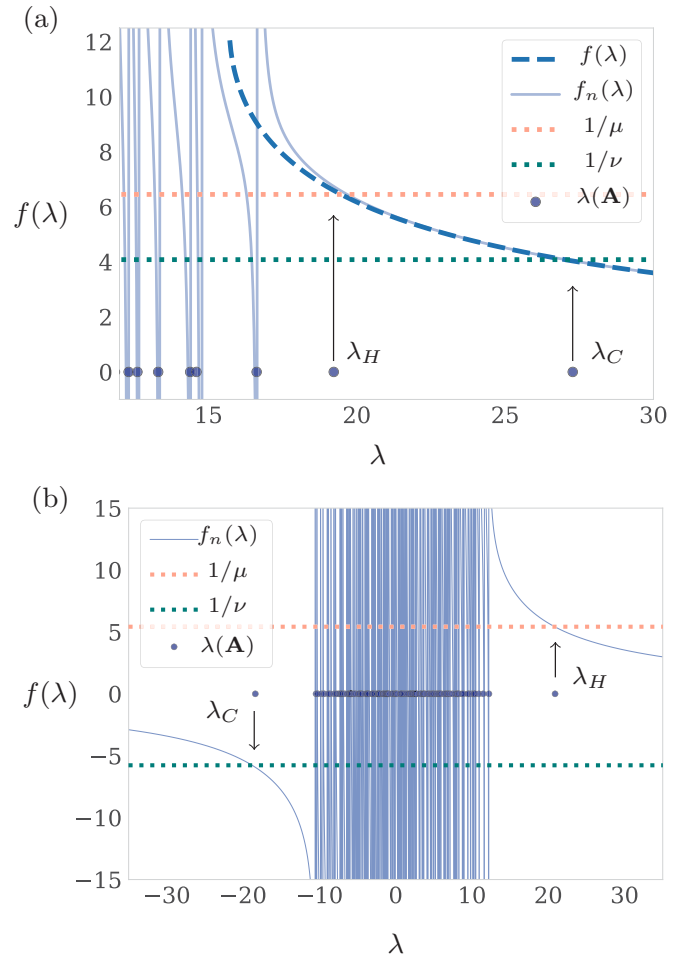


FIG. 11. Intersections of the function $f(\lambda)$ and $1/\mu$ and $1/\nu$ yield eigenvalue locations. (a) $f_n(\lambda)$ is the numerical solution to $f(\lambda)$ and intersects with $1/\mu$ and $1/\nu$ at λ_H and λ_C . $f(\lambda)$ is only defined away from the spectral band; the analytical solution diverges from the numerical approximation upon approaching the spectral edge. (b) The case where λ_H and λ_C are on opposite sides of the spectral band as the signal eigenvalues can be either positive or negative in signed networks.

Now we have a closed form solution for the Stieltjes transform of the spectral density of $\mathbf{X}$ which gives us an analytical formula for $f(\lambda)$ in the region when λ is outside of the spectral band, $|\lambda| > 2\sigma\sqrt{N} = \gamma$,

$$f(\lambda) = \text{Tr}(\lambda - \mathbf{X})^{-1} = \frac{\lambda \pm \sqrt{\lambda^2 - 4N\sigma^2}}{2\sigma^2}, \quad |\lambda| > \gamma. \quad (\text{B26})$$

We now solve for the eigenvalue λ of $\mathbf{A}$ that corresponds to the contrast signal,

$$f(\lambda) = \frac{1}{v}, \quad (\text{B27})$$

$$\frac{\lambda \pm \sqrt{\lambda^2 - 4N\sigma^2}}{2\sigma^2} = \frac{1}{v} \quad (\text{B28})$$

$$\Rightarrow \lambda = vN + \frac{\sigma^2}{v}, \quad |v| \geq \frac{\sigma}{\sqrt{N}}. \quad (\text{B29})$$

If $|v| \geq \frac{\sigma}{\sqrt{N}}$, then $|\lambda| \geq |\gamma|$, and the contrast eigenvalue λ is outside the spectral band, otherwise the leading eigenvalue is included in the spectral band. The analogous formula for the eigenvalue corresponding to the homogeneous signal is found by replacing v by μ in Eq. (B29). Figure 11 shows how the intersections between $f(\lambda)$ and $1/\mu$ and $1/v$ generate the eigenvalue locations. We have therefore found the same solutions for the signal eigenvalues of $\mathbf{A}$ using random matrix theory and complex analysis as done via perturbation theory in the main text, Eqs. (57) and (58).

-
- [1] A. Kirkley, G. T. Cantwell, and M. E. J. Newman, *Phys. Rev. E* **99**, 012320 (2019).
 - [2] J. Bruggeman, V. A. Traag, and J. Uitermark, *Amer. Sociol. Rev.* **77**, 1050 (2012).
 - [3] G. Facchetti, G. Iacono, and C. Altafini, *Proc. Nat. Acad. Sci. USA* **108**, 20953 (2011).
 - [4] Z. Maoz, L. G. Terris, R. D. Kuperman, and I. Talmud, *J. Politics* **69**, 100 (2007).
 - [5] P. Doreian and A. Mrvar, *J. Soc. Struct.* **16**, 1 (2015).
 - [6] S. Fortunato and D. Hric, *Phys. Rep.* **659**, 1 (2016).
 - [7] M. E. J. Newman, *Phys. Rev. E* **74**, 036104 (2006).
 - [8] S. Chauhan, M. Girvan, and E. Ott, *Phys. Rev. E* **80**, 056114 (2009).
 - [9] V. A. Traag and J. Bruggeman, *Phys. Rev. E* **80**, 036115 (2009).
 - [10] J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E. De Luca, and S. Albayrak, in *Proceedings of the SIAM International Conference on Data Mining* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 2010), pp. 559–570.
 - [11] P. Esmailian and M. Jalili, *Sci. Rep.* **5**, 14339 (2015).
 - [12] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, *Phys. Rev. Lett.* **107**, 065701 (2011).
 - [13] R. R. Nadakuditi and M. E. J. Newman, *Phys. Rev. Lett.* **108**, 188701 (2012).
 - [14] X. Zhang, R. R. Nadakuditi, and M. E. J. Newman, *Phys. Rev. E* **89**, 042816 (2014).
 - [15] A. Ghasemian, P. Zhang, A. Clauset, C. Moore, and L. Peel, *Phys. Rev. X* **6**, 031005 (2016).
 - [16] M. Wilinski, P. Mazzarisi, D. Tantari, and F. Lillo, *Phys. Rev. E* **99**, 042310 (2019).
 - [17] E. Abbe, *J. Mach. Learn. Res.* **18**, 6446 (2017).
 - [18] K. Kulakowski, P. Gawroński, and P. Gronek, *Int. J. Mod. Phys. C* **16**, 707 (2005).
 - [19] S. A. Marvel, J. Kleinberg, R. D. Kleinberg, and S. H. Strogatz, *Proc. Natl. Acad. Sci. USA* **108**, 1771 (2011).
 - [20] M. Girvan and M. E. J. Newman, *Proc. Natl. Acad. Sci. USA* **99**, 7821 (2002).
 - [21] M. Newman, *Networks*, 2nd ed. (Oxford University Press, Oxford/New York, 2018).
 - [22] B. Karrer and M. E. J. Newman, *Phys. Rev. E* **83**, 016107 (2011).
 - [23] P. Erdős and A. Rényi, *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17 (1960).
 - [24] M. E. J. Newman, *Phys. Rev. E* **67**, 026126 (2003).
 - [25] C. Sarkar and S. Jalan, *Chaos* **28**, 102101 (2018).
 - [26] D. Cartwright and F. Harary, *Psychol. Rev.* **63**, 277 (1956).
 - [27] V. Traag, P. Doreian, and A. Mrvar, Partitioning signed networks, in *Advances in Network Clustering and Blockmodeling*, edited by P. Doreian, V. Batagelj, and A. Ferligoj (Wiley Online Library, 2019), pp. 225–249.
 - [28] V. A. Traag, P. V. Dooren, and P. D. Leenheer, *PLoS ONE* **8**, e60063 (2013).
 - [29] R. Brubaker and D. D. Laitin, *Annu. Rev. Sociol.* **24**, 423 (1998).
 - [30] S. N. Kalyvas and M. A. Kocher, *Politics Soc.* **35**, 183 (2007).
 - [31] L.-E. Cederman and M. Vogt, *J. Conflict Resolut.* **61**, 1992 (2017).
 - [32] https://github.com/mmtree/Community_signed_networks.
 - [33] E. P. Wigner, *Ann. Math.* **67**, 325 (1958).
 - [34] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, Springer Series in Statistics (Springer, New York, NY, 2010).
 - [35] Z. D. Bai and Y. Q. Yin, *Ann. Prob.* **16**, 863 (1988).
 - [36] L. Erdős, *Russian Math. Surveys* **66**, 507 (2011).
 - [37] Y.-T. Li, D.-Z. Liu, and Z.-D. Wang, *J. Theor. Prob.* **24**, 1063 (2011).
 - [38] R. Basu, A. Bose, S. Ganguly, and R. S. Hazra, *Stat. Prob. Lett.* **82**, 1430 (2012).
 - [39] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins Studies in the Mathematical Sciences (Johns Hopkins University Press, Baltimore, MD, 1996).

Appendix 5

Militants and Mixed Messages: The Effect of Sponsor Ambiguity on Insurgent Ideology and Networks

Militants and Mixed Messages: The Effect of Sponsor Ambiguity on Insurgent Ideology and Networks

Nora Webb Williams

University of Illinois

Calvin Garner

University of Washington

Michael Gabbay

University of Washington

Abstract

While material resources have been the dominant focus of research on drivers of factional dynamics within insurgent movements, recent studies have shown that ideology also plays an important role. We address the interaction of material and ideological factors by theorizing and empirically investigating how state sponsorship impacts ideological diversity and patterns of cooperation and conflict in insurgent movements. When a movement relies upon a single state sponsor, conformance to the sponsor's preferences might be thought to render ideology irrelevant. However, we argue that when the state sponsor's position on an ideological dimension is ambiguous, then militant cooperation and conflict will be structured by ideological similarity (homophily) on that dimension. Conversely, sponsor clarity inhibits homophily. We test our theory on the Russia-backed insurgency in eastern Ukraine using an innovative empirical and methodological approach that combines an event and ideology database built from news articles and social media posts with signed network analysis. By analyzing an ideological dimension for which Russia's position was ambiguous and one for which it was clear, we find support for the hypothesized homophily-promoting and inhibiting effects of sponsor ambiguity and clarity, respectively.

Introduction

State sponsors provide resources to insurgent groups, such as arms, funds, training, and safe havens, that greatly enhance their military capabilities. Scholars have shown that these resources shape militant infighting dynamics as material concerns affect patterns of conflict and cooperation (Bapat & Bond, 2012; Tamm, 2016; Popovic, 2018; Bakke, Cunningham, & Seymour, 2012; Christia, 2012). Yet we still know little about how sponsor ideology affects militant dynamics, despite recent work emphasizing the importance of ideology in understanding fragmented civil conflicts (Gade, Hafez, & Gabbay, 2019; Gade, Gabbay, Hafez, & Kelly, 2019). In conflicts where external support is pivotal, analysis guided by prevailing rationalist and materialist perspectives on insurgent behavior would overlook the potential importance of ideology for intergroup dynamics. Such thinking would hold that militant factionalism is driven by competition for the material advantages offered by state sponsor support. Ideology would at best serve as a rhetorical cloak of little import to insurgent decision making. Yet, in this article, we show that ideology does indeed shape militant factionalism in a sponsorship context where it would be least expected to do so — an insurgent movement critically dependent upon the support of a single state sponsor. In fact, for the case we consider, Eastern Ukraine, the dependence upon Russia was so deep that in addition to the flow of arms and money, key leaders of the insurgency had backgrounds in Russian intelligence services and, ultimately, Russia's direct military intervention was required to stave off the movement's defeat.

The single state sponsor context we treat here is a particular case within the broader research question of how state sponsorship and ideology interact to shape the dynamics between militant groups. The single state case is a hard test of whether and how ideology matters for fragmented insurgent movements (i.e. movements with multiple, independent groups). For ideology to matter in this case would be puzzling from the power-centric orientation that dominates research on fragmented movements. Important works emphasize the distribution of power and its implications for organizational survival (Christia, 2012; Bakke et al., 2012; Krause, 2017). This power-centric focus also is prevalent in works on state sponsorship in fragmented movements, which examine the ability of a state sponsor to selectively reward and punish insurgent factions in order to induce their cooperation or perhaps encourage fragmentation (Bapat & Bond, 2012; Tamm, 2016; Popovic, 2018).

Consistent with the materialist perspective on intrastate conflict, external sponsor-

ship has been shown to be a crucial factor in insurgencies and civil wars, extending their duration and increasing the likelihood of conflict recurrence (Byman, Chalk, Hoffman, Rosenau, & Brannan, 2001; Brandt, Mason, Gurses, Petrovsky, & Radin, 2008; Cunningham, 2010; Regan, 2002; Sawyer, Cunningham, & Reed, 2017; San Akca, 2016; Karlén, 2017). However, a growing body of scholarship has also shown the importance of ideology to a range of militant behaviors such as recruitment and attacks against civilians (Costalli & Ruggeri, 2015; Hirose, Imai, & Lyall, 2017; Leader Maynard, 2019). For fragmented movements in particular, recent work has shown that ideology shaped networks of cooperation and conflict among insurgent groups in the Syrian civil war (Gade, Gabbay, et al., 2019; Gade, Hafez, & Gabbay, 2019).

In this article, we present a theory for whether and how ideology structures patterns of insurgent conflict and cooperation when state sponsorship is crucial for group survival. We then test the theory for the case of a single state sponsor, the situation in which ideology would be thought least likely to matter. In this case, logic suggests that ideology should not shape the militant network as the only viable militant groups should be those whose stated ideology matches the positions of the external sponsor, therefore garnering material support and living to fight another day against either the state or their fellow rebels. However, we theorize that there is still room for a window of ideological variation in the single state sponsor case. We argue that, as ideology is multidimensional, if there are dimensions of ideology along which the external state sponsor is ambiguous, militant groups can differentiate themselves along those lines while still receiving external material support. Those ideological differences then shape patterns of intra-movement cooperation and conflict.

This paper makes theoretical, empirical, and methodological contributions to the literature on insurgencies and civil wars. Our theoretical explanation for how ideology can impact insurgent factional dynamics in the presence of an external state sponsor integrates ideological homophily, the disaggregation of ideology into separate components, and state sponsor ambiguity regarding an ideological component. Ideological homophily implies that ideologically similar groups will have higher rates of cooperation whereas ideologically dissimilar groups will have higher rates of conflict. The disaggregation of ideology considers how militants frame the conflict (which has been shown to be mobilizing in other post-Soviet conflicts; see Shesterinina (2016)), their idealized post-conflict

order for state and society, and their desired territorial domain. We argue that not all components of militant ideology will matter all the time. We theorize that, in the presence of a single external sponsor, the components of ideology about which the sponsor is ambiguous will matter most in structuring patterns of conflict and cooperation. When an external sponsor is clear about a component of militant ideology, then we expect that component will not generate homophily.

Empirically, we test our hypotheses using a network analysis of media reports about the conflict in Eastern Ukraine, with data from February 2014 to August 2017. We code these media reports for episodes of conflict or cooperation between militant actors, which we use to construct the network, as well as statements of ideology by the actors, which we use to code actor positions on the ideological components. The anti-government insurgents were actively supported from the early stages of the conflict by the Russian government, whose position was clear on the dimensions of ideology involving framing the conflict and sociopolitical order. But Russia was initially ambiguous about its designs concerning the territorial status of the regions under contention.

Methodologically, we use an innovative framework based upon signed networks. Signed networks provide an integrated representation of cooperation and conflict that has not previously been employed to analyze insurgent movements. Community structure algorithms are used to reveal clusters in networks, which for the signed network case are marked by increased intracommunity cooperation and increased intercommunity conflict. This structure can then be tested for significant statistical relationships with the ideological components. Additionally, we employ an analysis based on the network concept of assortativity, a measure of homophily. Random permutations of the signs of existing network ties produce a null distribution for comparison with the observed assortativity. The application of these community structure and assortativity analyses to signed networks is, to our knowledge, a novel contribution to the political science literature in general. Finally, we also apply exponential random graph modeling to the separate positive and negative tie networks.

We test for homophily in the Ukraine case along two ideological dimensions, one in which the state sponsor's position was ambiguous and one for which it was clear. The ambiguous dimension involved militant territorial aspirations, specifically the annexation or independence of Novorossiia, an idealized, largely Russian-speaking region that some

insurgents sought to break away from the Ukrainian state. We find that clustering in the combined cooperation-conflict network is indeed predicted by militant proximity with respect to the level of support for Novorossiia. This finding not only supports our theory but also, more generally, further demonstrates ideology's importance to insurgent factionalism.

The ideological dimension on which Russia expressed a clear position was its conflict frame of "Russia vs. the West". Our theoretical prediction that homophily will not be present along this unambiguous component is borne out empirically. Surprisingly, however, not only is homophily absent but its opposite is present, heterophily, a tendency for militants who are more distant along the conflict frame dimension to have better relationships. While this heterophily finding goes beyond our theoretical expectations, it is consistent with the theory's characterization of ideological components with clear sponsor positions as being rhetorical in nature as opposed to the actionable nature of sponsor-ambiguous components.

We proceed by first presenting our theory, which also encompasses cases when multiple state sponsors are present. We then discuss the insurgency in eastern Ukraine, focusing on militant and Russian ideological positions. Next, we describe the data and variables we use to construct militant network ties and ideological positions. Our community structure and assortativity analysis methods are then presented, followed by the empirical findings resulting from their application. The conclusion includes a sketch of potential similarities between Russia's support for eastern Ukrainian insurgents with the United States' support for the Contra rebels in Nicaragua in the 1980s.

Theorizing Sponsorship, Ideology, and Militant Structure

External sponsors, particularly states, can dramatically improve the fighting ability of an insurgent group, and so it is not surprising that their importance to insurgent movements is well established in the literature. Recent work on this topic focuses on how an external sponsor may shape the behavior of an insurgent group, or why sponsors choose to back one group out of many possible options. Tamm (2016) argues that an external sponsor shapes behavior with the implicit threat of denying resources to an insurgent group it supports;

it can always punish undesirable behavior by insurgent leaders by re-allocating resources to the leaders' rivals within the insurgent movement or the group itself. This behavior is likely more pronounced when the movement has only one sponsor who is therefore free to search for and reward its ideal insurgent partners. Indeed, Sinno (2008, p. 248) argues that a single external sponsor will lead to a centrally controlled insurgent movement, while sponsorship from multiple sponsors is valuable to insurgent groups because it "shields the beneficiary organization from the abrupt cessation of aid if one sponsor decides to withhold support." Reaching a similar conclusion about multiple sponsors, Lidow (2016) argues that the presence of more than one external sponsor will increase the bargaining power of an insurgent group, which may explain why in Syria, a conflict with many external sponsors, sharing a state sponsor did not appear to make groups less prone to fight each other (Gade, Hafez, & Gabbay, 2019). This suggests that the capacity for a state sponsor to control its client groups is substantially diminished when other sponsors are present.

Material factors clearly matter in explaining intergroup dynamics in insurgencies such as infighting, cooperation, alliances, fragmentation, and organizational structure of groups within insurgent movements (Bakke et al., 2012; Christia, 2012; Krause, 2014; Staniland, 2014; Fjelde & Nilsson, 2018). Recent scholarship, however, has begun to temper this power-centric perspective, illuminating the important role played by ideology. Leader Maynard (2019) outlines the microfoundations to understanding how ideology and strategic interest overlap and work together to influence militant behavior. Other recent studies have shown the importance of ideology for explaining how group founders motivate fighters and supporters (Gutiérrez Sanín & Wood, 2014); whether fighters join a rebel group (Costalli & Ruggeri, 2015); the socialization of rebel group members (Ugarriza & Craig, 2013); how groups maintain cohesion (Oppenheim, Steele, Vargas, & Weintraub, 2015); and the distribution of violence against civilians (Hirose et al., 2017). What extant work has not yet resolved, however, is the effect of external state sponsorship on militant ideology and how it impacts group interactions.

We argue that ideological considerations play a role in external sponsors' decisions of which groups within a broad insurgent movement to support. External sponsors will attempt to back the group that best matches their preferences across ideology. In turn, militants will also seek out ideologically compatible sponsors. Consequently, the range of sponsor and militant ideological positions will determine the extent to which ideological

diversity can flourish. This diversity then provides space for ideological homophily to shape networks of insurgent behavior; those who are most similar in terms of ideology will cooperate with each other, while fighting against those who are ideologically dissimilar.

It is important to recognize that ideology is multidimensional and that some dimensions may exhibit homophily while others do not. We employ the framework of Gade, Gabbay, et al. (2019), which highlights three conceptually distinct components of insurgent ideology: (1) *conflict frame*, corresponding to the primary division into in- and out-groups in militant discourse; (2) *ideal polity*, the vision of sociopolitical order espoused by militants; and (3) *territorial aspiration*, the extent to which the domain militants seek to govern severs the territorial integrity of the state.

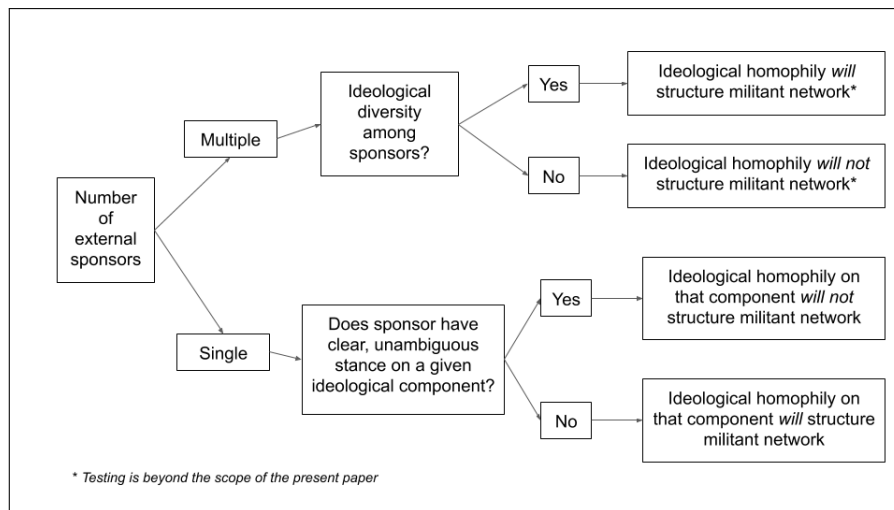


Figure 1: Theory of external sponsorship, ideology, and militant network structure

Our full theory considers the effect of the number of state sponsors and the diversity and clarity of their ideological positions on militant network homophily along ideological components. Figure 1 presents the theory graphically. We assume that state sponsorship is essential to the viability of insurgent groups (either in the fight against counterinsurgents or in competition with fellow militants who receive sponsor support), and so their leaders must take sponsor preferences into account. When there are multiple external state sponsors that disagree on component(s) of insurgent ideology, insurgents can either seek out sponsors with kindred ideological positions, or, more pragmatically, strategically position themselves as ideological partners of the sponsor. As the number of external sponsors and the disagreement among them grows, the potential for ideologically diversity to shape cooperation and conflict within a given anti-state movement also grows. An insurgency is most likely to be comprised of numerous, ideologically varied, powerful groups if

the insurgency has the backing of many different external sponsors that exhibit a great deal of variation in their own ideologies. In this situation, we should most expect to see ideological homophily shaping patterns of insurgent cooperation and conflict. The Syrian civil war serves as evidence in support of this prediction: there were at least five state sponsors of the opposition to the Assad regime, an opposition marked by both ideological diversity and attendant homophily (Gade, Hafez, & Gabbay, 2019; Gade, Gabbay, et al., 2019). In cases where the multiple sponsors have similar and clear ideological positions, we do not expect to observe homophily in the militant network.

But what about when there is only one external sponsor? First blush intuition would hold this as a hard case for ideology to affect conflict and cooperation; competition between militant groups along ideological lines should not be relevant as there is only one “correct” ideology that would earn external sponsor support. Indeed, we theorize that this intuition is warranted but with the caveat that the sponsor’s ideological position is clear; given strong sponsor preferences, only groups operating in a narrow ideological space will thrive. This situation is in line with the expectation from prior work that in a conflict with one external sponsor, that sponsor is free to reward and punish behavior by its client groups, ensuring behavior that is consistent with its preferences (Tamm, 2016; Sinno, 2008; Lidow, 2016; Salehyan, Gleditsch, & Cunningham, 2011). Consequently, a single external sponsor, demanding adherence along a given ideological component, will limit the possibility for this component to explain inter-group dynamics and generate homophily.

Furthermore, a sponsor’s clear position renders an ideological component largely rhetorical: even if its militant clients espouse deviations from this position, the sponsor is unlikely to allow any goals and other tangible implications associated with those deviations from being enacted on the ground if they are inconsistent with the sponsor’s line. The rhetorical nature of the component dampens the impetus for militant cooperation to be organized along ideological lines to reach shared goals. This further inhibits homophily beyond the ideological narrowing caused by the pressure to parrot the sponsor’s line.

The absence of homophily outcome is not inevitable, however, because external sponsors do not always have clear positions on all components of ideology. The existence of ideological dimensions along which the external sponsor is ambiguous opens up the potential for militant differentiation along those dimensions. An external sponsor may be genuinely undecided about a component due to ambivalence or internal factionalism.

Alternatively, it may deliberately project ambiguity for strategic reasons, creating useful uncertainty among its opponents. Regardless of the reason, this ambiguity will enable insurgent groups to embody different positions on the component. The differentiation allows militant groups to appeal to distinct pools of fighters, bases of popular support, or factions within the external state sponsor. Moreover, unlike sponsor clarity which makes ideological differences moot, sponsor ambiguity implies that differences in militant ideological positions are actionable in terms of ends and means. As a result, sponsor ambiguity not only opens up more space for insurgents to air their ideological differences, it also makes possible on-the-ground realization of these differences. The combined effect of a larger as well as actionable window of disagreement facilitates the operation of homophily along the sponsor-ambiguous component, encouraging joint action among ideological siblings to achieve similar goals and limiting counterproductive infighting. Conversely, insurgents with discordant ideological positions will be less likely to work together and more likely to clash.

Summarizing the above arguments leads to the two hypotheses, one positive and one negative with respect to homophily, that we test in this paper:

- **H1:** If the sole external sponsor is ambiguous about a given component of ideology, then militant cooperation and conflict will exhibit homophily along that component.
- **H2:** If the sole external sponsor is clear about a given component of ideology, then patterns of cooperation and conflict will not exhibit homophily along that component.

The Case: Conflict in Eastern Ukraine

Given the presence of a sole external sponsor, Ukraine is a hard case of the more general theory presented in Figure 1. In Ukraine, we observe variation in the militant groups along ideological dimensions and evidence of conflict and cooperation between groups. Scholars note local recruitment of fighters based on ideological appeals, as well as contractual and material inducements (Kudelia, 2019). The case has a single-state sponsor, Russia, whose support was essential to the viability of the insurgency as a whole and for individual groups. The case also allows for both of the hypotheses stated above to be tested because Russia's position was ambiguous on one component of ideology but not on the others.

Case Background

The conflict in Eastern Ukraine began in the spring of 2014, after a six-month period of almost continuous upheaval in Ukrainian politics, including the Maidan protests in Kyiv, the flight of then-President Viktor Yanukovych to Russia, and the Russian invasion of Crimea. What began as protests around government buildings in Donetsk and Luhansk oblasts (collectively referred to as the Donbas region) soon became an armed conflict, with anti-government fighters hoping to gain increased autonomy from Kyiv, to secede from Ukraine and become independent, or to join Russia. The conflict, especially the early phase, was characterized by numerous quasi-independent groups operating against Ukrainian military and volunteer forces. Many of the early leaders were originally from Russia, some with connections to Russian intelligence and security forces.

Insurgents eventually consolidated control in the eastern regions of Donetsk and Luhansk oblasts, though the militant groups remained markedly personalized. Two political units, the Donetsk People's Republic (DNR) and the Luhansk People's Republic (LNR),¹ were established after controversial referenda in May 2014 and are not recognized as sovereign states on the world stage. The LNR and DNR established legislative bodies, developed official military hierarchies, and held elections.

Russia's support for the insurgents is a clear factor in the ability of the LNR and DNR to continue to resist attempts by the Ukrainian government to re-integrate the two territories. Russia has provided military, financial, and political support, such as recognizing the 2014 referenda in the regions (Williams, 2014; Gall, 2018; Kramer & Gordon, 2014), providing advanced weaponry including the system used to shoot down Malaysia Airlines Flight 17 in July 2014 (Zaken, 2018), and sending troops to stop a Ukrainian advance in late summer 2014 (International Crisis Group, 2014, pp. 2-4). Insurgents were heavily dependent on Russian military, political, and logistical support, putting the Russian government in the position to heavily influence both the behavior and the ideological line of the insurgents.

Fighting between government forces and insurgents continued throughout summer 2014. On September 5, 2014, an initial ceasefire was agreed to, now referred to as Minsk I. A second ceasefire, Minsk II, was signed on February 12, 2015. A "line of contact" between the belligerents was established and more or less respected, although periodic

¹The abbreviations DNR and LNR are based on the transliteration of the Russian-language acronyms.

violations of the ceasefire in the form of exchanges of fire continued throughout the data collection period.

Novorossiya as Insurgent Ideology

At various times and to various degrees, leaders in the DNR and LNR have used the word Novorossiya (New Russia) to describe a project that would combine their territorial units into one structure and dramatically expand territory controlled by the insurgents to include most of southeast Ukraine. Some individual insurgent leaders suggested that Novorossiya would eventually include much larger territory, or a federation of up to eight separate states. Attempts were made to combine insurgent combat forces into one body, the Novorossiya Armed Forces.

In our reading of ideological statements by insurgent actors, the main emphasis of the Novorossiya project was on the extent of territory that would be administered by non-Ukrainian powers (whether independently or as a part of Russia). Thus, statements in favor or against Novorossiya often relate to territorial ideological claims.

Analysts of the Ukrainian conflict identify differences regarding Novorossiya as a main cleavage within the insurgent movement (International Crisis Group, 2014) and also among the local population (O'Loughlin, Toal, & Kolosov, 2017). The International Crisis Group (2014) notes a split in 2014 between the pragmatists and hardliners, with the hardliners prepared to fight for an expanded Novorossiya and pragmatists willing to accept a slow, more peaceful integration into Russia or some independent configuration. The 2014 ICG report identifies Igor Strelkov, Pavel Gubarev, and Alexei Mozgovoi as hardliners. Insurgent government insiders, such as Alexander Zakharchenko, Andrei Purgin, and Denis Pushilin serve as examples of pragmatists.

Russian Ideological Positions

Along two of the key components of ideology—conflict frame and ideal polity—we see a clear position on Russia's part that signaled to insurgents the boundaries of acceptable positioning. With respect to conflict frame, Russia situated the fighting in Eastern Ukraine as part of its ongoing and escalating tensions with the West, broadly defined, including the United States, the European Union, and NATO. Russia defined the conflict in Ukraine as one in which the West was inserting itself where it did not belong, casting the Maidan

uprising as a Western-sponsored coup. The U.S. and E.U. objected to Russian military action in Ukraine, ultimately imposing heavy sanctions, and Russia responded with anti-Western rhetoric. Anti-Western narratives emanating from the Russian government and large Russian media companies are nothing new, but these narratives gained “particular vitriol” when framing the conflict in Eastern Ukraine (Hutchings & Szostek, 2016, p. 174). The objective of these narratives was to blunt and undermine U.S. criticism of Russian action in Ukraine, ascribing any criticisms to a Cold-War mentality in the West (Ambrosio, 2016). It would have been nearly impossible for anyone consuming Russian media during this period to miss this narrative; the Russian government position on conflict frame was obvious.

While the Russian government position on ideal polity was not as clearly stated as its position on the conflict frame, its position on this component of ideology was neither ambiguous nor ambivalent to insurgent fighters in Eastern Ukraine. If Novorossiya were to be annexed by Russia, visions of the restoration of either the Romanovs or the commissars would be distant dreams requiring the fall of Putin’s existing regime. If Novorossiya were to be independent, insurgents could look to examples of other regimes that Russia has supported in the post-Soviet space, namely South Ossetia, Abkhazia, and Transnistria. In all of these cases there are nominal democratic institutions that exist alongside corruption and limited democratic choice. In confirmation of the lack of viable options in terms of ideal polity, the LNR and DNR governments did indeed hold elections of questionable openness to fill parliaments of questionable power, much as in Russia.

Unlike conflict frame and ideal polity, the Russian government has sent ambiguous signals about the territorial aspiration that it would support, especially in the early days of the conflict (MacFarquhar, 2014; “East Ukraine separatists seek union with Russia”, 2014; Bowen, 2019; International Crisis Group, 2019). One way to track Russia’s ambiguity concerning territorial aspiration is to see how its position on the concept of Novorossiya has changed over time. Table 1 contains notable quotations about the Novorossiya project from both Russian political leaders and insurgents. Prior to May 2015, there were some encouraging signals from Russian officials about support for an independent Novorossiya, including an extended comment from Putin on the historical “correctness” of the endeavor. However, Russia tempered such comments with signals indicating less-than-wholehearted support for the Novorossiya project. Politically, the Russian reaction to the indepen-

dence referenda in the LNR and DNR fell short of full endorsement (MacFarquhar, 2014). Militarily, support on the ground was mixed for further territorial expansion and fell far short of full-scale invasion and occupation that occurred in Crimea. Notably, insurgent leader Igor Strelkov made comments criticizing Russia for refusing to offer more military support, which he seemed to believe would have helped avoid an embarrassing retreat from Ukrainian government forces. By May 2015, there were strong indications that the Novorossiya project was dead – Russia had ceased to be ambiguous about this component of ideology and had decided that the territorial provisions in Minsk II were acceptable. The two quotes from May 2015 in Table 1 demonstrate the Russian state decision, as part of Minsk II, that the territorial aspirations for Novorossiya no longer had Russian support and the acknowledgement from insurgents that the support for Minsk II meant the “freezing” of Novorossiya.

Table 1: Timeline of Support for Novorossiya from Russia

Date	Quote	Source	Position
2014-04-17	I would like to remind you that what was called Novorossiya (New Russia) back in the tsarist days – Kharkov, Lugansk, Donetsk, Kherson, Nikolayev and Odessa – were not part of Ukraine back then. These territories were given to Ukraine in the 1920s by the Soviet government. Why? Who knows... Russia lost these territories for various reasons, but the people remained... we ought to do everything we can to help these people defend their rights and determine their fate on their own. This is what we will fight for. Let me remind you that the Federation Council of Russia gave the President the right to use the Armed Forces in Ukraine.	Russian President Vladimir Putin	Encouraging statements about Novorossiya
2014-05-06	Favorable discussion of Federal Republic of Novorossiya	Former Ukrainian MP and Novorossiya leader Oleg Tsarev	Supportive of Novorossiya
2014-05-22	Proclaimed foundation of Novorossiya	DNR ‘People’s Governor’ Pavel Gubarev	Proclaimed foundation of Novorossiya
2015-05-19	“At all levels, through the mouth of the Russian president, in other formats we say that we want [DNR and LNR] to become part of Ukraine. They have now submitted their draft constitution. There they are talking exactly about the status provided for by the Minsk agreements: the republics will be part of Ukraine and the constitutional reform will continue to be consolidated on A permanent basis.”	Russian Minister Lavrov Foreign Minister Sergey	Statements completely incompatible with Novorossiya
2015-05-20	“The activity of Novorossia’s structures is frozen because... its existence contradicts the Minsk agreements.”	Former Ukrainian MP and Novorossiya leader Oleg Tsarev	Suspending official Novorossiya movement

According to our theory, the ideological dimension(s) along which the external state sponsor is ambiguous should induce homophily in the patterns of conflict and cooperation between militant actors. In the case of the conflict in eastern Ukraine, Russia was initially ambiguous about the territorial aspiration component, as expressed in statements about

Novorossiia. Conversely, Russia's clarity regarding conflict frame and ideal polity implies that these components should not leave homophily signatures in the conflict-cooperation network.

Data and Sources

To test the theory, we identified relevant actors, identified cooperative and conflictual events between these actors, and built measures of the actors' positions on relevant components of ideology. Our data sources were newspaper articles, media digests, and social media posts translated and made available by the U.S. government.² We then used a manual coding scheme to extract relevant events and ideological positions from a subset of the documents.

Based on named entity analysis of the entire translated document corpus, we identified documents with a high density of relevant actor names in the text. We then hand-coded 136 documents, which we believe account for roughly 20% of the total mentions of the key actors that are identified in our dataset, although this is likely a conservative estimate. Many documents were pulled into our dataset because the name of insurgent groups of actors were either common surnames or identical to place names. The Dnipro Battalion, for example, shares its name with the Dnipro River and the city Dnipro, capital of Dnipropetrovsk oblast. Similarly, some individuals who may have been connected to the conflict in some capacity, such as Ukrainian oligarch Rinat Akhmetov, were also frequently included in articles that are irrelevant to the conflict. Thus, we believe that the documents we coded represented an even higher percentage of relevant actor mentions. We next briefly describe how we determined our set of relevant actors, labeled events and event types, and our strategy for assigning actor ideology. A full discussion of our coding methodology is available in Appendices A and B.

Actors

The universe of potentially relevant actors involved in the Ukrainian conflict is vast, with militia groups and commanders on both sides, as well as oligarchs and politicians with more murky ties to combat groups. For the purposes of this analysis, we focus on

²In total, we collected 3,528 unique documents that contained at least one search term of interest from January 3, 2014 to August 1, 2017.

a subset of all actors: insurgent leaders. We focus on leaders, as opposed to specific battalions, because individuals were more often observed making ideological statements in this context (that is, we see more impassioned speeches on YouTube from individual commanders than manifestos for particular battalions).³ Appendix A details our process of identifying leaders. In total, we collected data on 28 unique insurgent leaders.

Event Types

An event in our database follows the structure of someone (the “from” actor) doing something (the event action) to someone else (the “to” actor). We only coded events where both the “from” and “to” actors were included in our list of actors to track. For each event, we also noted the time of the event and the published date of the source for the event.

When we encountered a report of cooperative or conflictual behavior, we coded as many different dyadic interactions as the event contained. Each dyad constitutes an event in our coding scheme, and each event was coded as either **conflict** or **cooperation**. For example, a report of two insurgents cooperating to attack another insurgent would contain three events: one cooperative and two conflictual.⁴ In total, the coding yielded 171 total events involving relevant actors, of which 86 were instances of cooperation and 85 were conflict.

Ideology

To construct a variable for conflict frame, we determined actor viewpoints, positive or negative, toward broad parties or social identities relevant to the conflict as either potential in-groups or out-groups. In particular, we coded actor statements with respect to the following conflict parties: Russian people; Russian state/military; Russian oligarchs; Ukrainian people; Ukrainian state/military; Ukrainian oligarchs; LNR/DNR people; LNR/DNR state/military; LNR/DNR oligarchs; the U.S./NATO/West; the Maidan movement; Cossacks identity, Orthodox Christianity identity; and an open-coding for

³Where relevant we consolidated groups to individuals. Thus, for example, all references to the Vostok battalion were ultimately assigned to Alexander Khodakovsky, the group’s leader. When the leadership was unclear, or when it shifted over time, we did not consolidate a group to an individual.

⁴Cooperative (1): insurgent A cooperates with insurgent B to attack insurgent C. Conflictual (2): insurgent A attacks insurgent C; insurgent B attacks insurgent C. We further classify an event subtype for each act of cooperation or conflict as either rhetorical, violent, material/military, or relational. For information on the distribution of subtypes, see Appendix B.

“other” as appropriate. When these statements were positive or negative, we coded that statement as indicative of an ideological position for that actor towards a given party to the conflict, placing the party on one side of an us/them, in-group/out-group divide. A negative statement resulted in a -1 score for an actor towards a conflict party, indicating that the conflict party was framed as an out-group, while a positive statement resulted in a +1 score, indicating that the conflict party was framed as an in-group.

Exploratory factor analysis and the case history led us to conclude that the main axis for the conflict frame was **Russia vs. West**. Insurgent actors typically made statements indicating their opposition to either Ukraine/the West (the Ukrainian state/military, the U.S./NATO/West, and the Maidan movement) or support for Russia (Russian state/military, Russian people, and the LNR and DNR). This conflict frame is in keeping with research that has analyzed Russian rhetoric in the years leading up to and after the outbreak of war in Eastern Ukraine.

To determine actors’ preferences with respect to ideal polity, we coded for statements regarding the ways in which they sought to structure the relationship between the state and society. Specifically, we sought to operationalize the differences between “reds,” “whites,” and “browns” that Laruelle (2016) identifies as indications of striving towards a communist, imperial, and fascist polities. However, we found little meaningful variation in this variable, and Laruelle acknowledges that, in fact, there is considerable overlap among some of the actors to whom she ascribes these categories.

We construct our measure of territorial aspiration from statements in support of or opposition to the idea that **Novorossiia** should be independent or become part of Russia. Examples in support of Novorossiia include Igor Strelkov’s statement that, “I hope that Novorossiia will make it... as a new state allied with Russia” and Pavel Gubarev’s claim that the issue was one of “a matter of life” for Strelkov and himself. Statements coded as opposition include Khodakovsky’s advocacy for the regions to be reintegrated into Ukraine, but with greater autonomy. Because the official justification for abandoning the pursuit of Novorossiia was framed as a logical extension of the Minsk Agreements (as done by one-time Novorossiia leader Oleg Tsarev, see Table 1), we also considered statements in opposition to the Minsk process as being fundamentally about wanting to continue the fight and having greater territorial ambitions. Thus anti- and pro-Minsk statements are also included in the Novorossiia variable.

Ideology statements were less frequent in the documents we encountered than either cooperation or conflict events between actors. Our coding yielded 261 ideological statements that were associated with a total of 49 actors that appear in our events database. In coding for ideology, actors were assigned a zero for any ideological component for which no relevant statements were identified. Actors with zeros for all ideological components were dropped from the analysis. We normalize the raw scores so that each actor's scores on conflict frame and territorial aspiration are between -1 and 1.

An initial review of the range of scores for Russia vs. West and for Novorossiia show that there was, in fact, more ideological variation on the territorial aspiration variable than on the conflict frame variable. The values assigned to anti-government actors or groups on the conflict frame variable ranges from 0 to 1, meaning that we recorded no incidents of anti-government insurgents supporting a conflict frame that identified with the Ukrainian state in opposition to the Russian state. Insurgent actors either promoted a pro-Russia conflict frame or kept quiet on this front. The territorial aspiration variable, on the other hand, saw actors both in support of and in opposition to the idea of fighting for Novorossiia as a contiguous territory. The scores for anti-government insurgents ranged from -1 to 1 on this component of ideology. The fact that there was more variation on the territorial aspiration variable than on the conflict frame variable is in keeping with our expectations based on the initial ambiguity of Russia, the sole external sponsor for the insurgency, on territorial aspiration.

Method

We test our hypotheses with social network analysis tools, which focus on tie formation (the presence of a given behavior, e.g., cooperation or conflict) between “nodes” (e.g., actors, individuals, groups, nations) and the predictors of those ties at the level of the node, dyad, or network. The application of social network analysis within political science and international relations has emerged over the past decade (Maoz, 2011; Minhas, Hoff, & Ward, 2016; Desmarais & Cranmer, 2017). It is an increasingly common strategy for analyzing militant dynamics given the proliferation of armed groups in civil wars (Metternich, Dorff, Gallop, Weschle, & Ward, 2013; Gade, Hafez, & Gabbay, 2019; Gade, Gabbay, et al., 2019; Dorff, Gallop, & Minhas, 2020).

Most network analyses in conflict studies focus on either the set of cooperative ties

or the set of conflictual ties, not both. In this study, we combine data on conflict and cooperation events to form a *signed network*, where cooperative ties are positive values and conflictual ties are negative. The complexity added to the data collection and analysis is warranted by the reality that decisions about cooperation and conflict are often inextricably linked. The adage “the enemy of my enemy is my friend” embodies this linkage and has been formalized in structural balance theory, the most prominent theory of signed social networks (Cartwright & Harary, 1956). Signed networks have not yet been widely adopted in political science although they have been used to investigate structural balance in the international system (Maoz, Terris, Kuperman, & Talmud, 2007).

To test the degree to which ideological homophily shapes the network, we employ community structure methods. Community structure refers to the tendency of networks to cluster into subgroups or “communities” where the nodes in each community have similar patterns of tie formation, both within and without the cluster. As clustering can arise due to preferential tie formation among nodes with like attributes, community structure is frequently generated by homophily processes. We use two methods for analyzing community structure: one is based on the “modularity” matrix whose elements assess the extent to which the observed dyadic tie value differs from that expected by chance; the other is based on the “adjacency” matrix whose elements are the dyadic tie values themselves. For our signed networks, the positive tie value is taken to be the number of cooperative events between the node pair and the negative tie value is the number of conflictual events (both taken without regard to the “from” and “to” direction). The signed adjacency matrix uses the net tie value, the difference between the numbers of cooperative and conflictual events in a dyad.

For simplicity, we describe the modularity matrix method in the context of unsigned networks, in which ties are only positive, and then we treat the signed case. For unsigned networks, the goal of community detection algorithms is to find subgroups of nodes with relatively dense ties within groups and sparse ties between groups. The modularity is a number that expresses the extent to which a candidate division of the network into different communities exhibits a level of intra-community linking exceeding that expected if the division were, in fact, arbitrary with no correspondence to behaviorally meaningful subgroups (Newman, 2006). The modularity matrix contains the potential contribution to the modularity from each dyad, realized if the nodes in the dyad are assigned to the

same community. The modularity matrix element for a dyad is the difference between their observed number of ties (that is, their adjacency matrix element) and the number of ties between them expected from a null model, which is taken to be proportional to the product of their respective degrees (total number of ties).⁵

The concept of modularity has been extended to signed networks (Traag & Bruggeman, 2009; Traag, Doreian, & Mrvar, 2020). Essentially, one seeks communities which have many positive and few negative ties within each community, with few positive ties and many negative ties between communities. The full modularity matrix is constructed as the difference between the separate positive-tie only and negative-tie only modularity matrices.⁶

The eigenvectors of the modularity matrix are useful for representing community structure (Newman, 2006). An eigenvector has a component for each node and distinct eigenvectors form independent directions which do not mix with each other under operation of the matrix.⁷ Eigenvalues are ranked from greatest to least. A single eigenvector (typically, but not necessarily, the first) can be used to divide nodes into two communities: any node whose component in the eigenvector is positive is assigned to one community while nodes with negative eigenvector components are assigned to the other.

Eigenvector components can be used to test for homophily without the explicit assignment of nodes to communities, as homophily is closely linked to modularity (Newman, 2006).⁸ If the homophily process is sufficiently strong, then it should be reflected in a highly-ranked eigenvector as indicated by a significant correlation between eigenvector and the node variable. Calculating the correlation with modularity matrix eigenvectors has previously been used to test the relationship between ideology and militant network structure in the unsigned case (Gade, Hafez, & Gabbay, 2019). We limit our analysis to the use of only the first two eigenvectors in our test for the presence of homophily among the ideology components in the next section. Although one could test correlations with each eigenvector in isolation, the optimal alignment of the variable with network

⁵Mathematically, we let the adjacency matrix elements A_{ij} be the number of ties between nodes i and j in an undirected network ($A_{ij} = A_{ji}$). The degree of a node is $k_i = \sum_{j=1}^N A_{ij}$ and the total number of ties is $m = \sum_i k_i$. The elements of the modularity matrix $\mathbf{B}$ are given by $B_{ij} = A_{ij} - k_i k_j / 2m$.

⁶Defining A_{ij}^+ , and A_{ij}^- to be the number of positive and negative ties between i and j respectively, then the signed network modularity matrix $\mathbf{B}$ is the difference of the positive and negative modularity matrices: $\mathbf{B} = \mathbf{B}^+ - \mathbf{B}^-$.

⁷Formally, the $(\lambda_k, \mathbf{u}_k)$ eigenvalue-eigenvector pairs of the modularity matrix satisfy $\mathbf{B}\mathbf{u}_k = \lambda_k \mathbf{u}_k$.

⁸For signed networks, homophily connotes the tendency for preferential positive (negative) tie formation with similar (dissimilar) nodes.

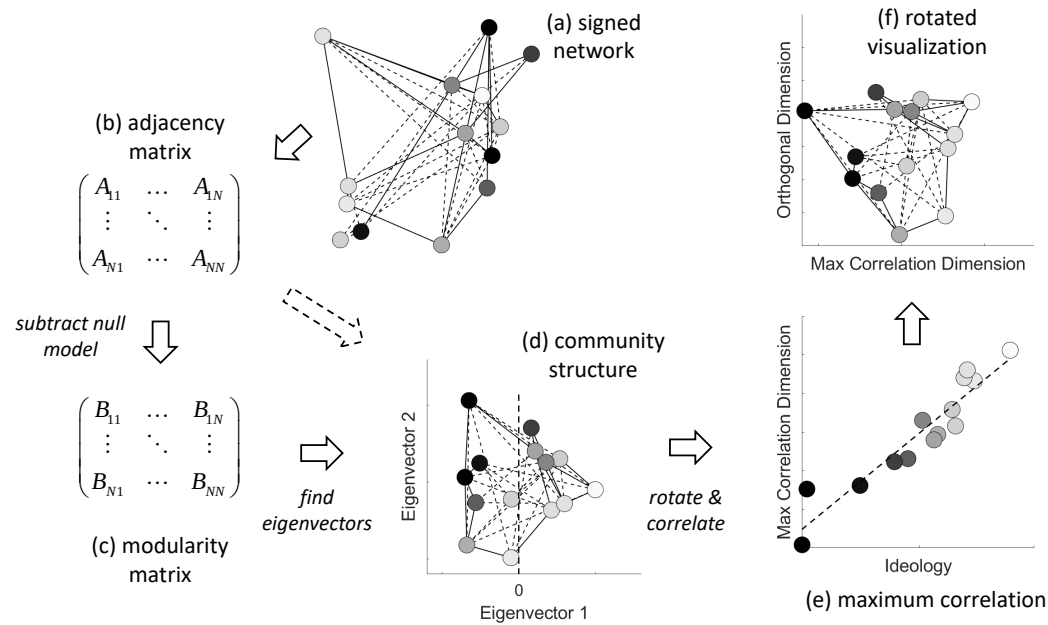


Figure 2: Overview of analysis procedure. (a) Signed network with N nodes (randomly placed). Cooperative (conflictual) ties shown as solid (dashed) links. Nodes shaded with respect to ideology component so that lighter nodes denote higher ideology values. (b) Adjacency matrix $\mathbf{A}$ whose elements are tie values between nodes. (c) Modularity matrix $\mathbf{B}$ formed by difference of positive and negative tie modularity elements in which null model is subtracted from observed tie values. (d) Eigenvector-based network visualization in which a node's coordinates are its components in the first two eigenvectors of $\mathbf{B}$ (the alternative adjacency-based method uses eigenvectors of $\mathbf{A}$ as indicated by dashed arrow). Community structure, given by the two communities of nodes left and right of the dashed vertical line, shows ideological clustering. (e) Eigenvector-based network is rotated to find the dimension along which the node positions maximally correlate with ideology. The homophily test is based on this correlation. (f) Rotated version of (d) showing the improved left-right ideological clustering obtained by using the maximum correlation dimension as the x-axis.

structure may not reside along a single eigenvector. To allow for this possibility, we will rotate the network in the two-dimensional space formed by these eigenvectors in order to find the maximum correlation. This rotation angle represents an additional parameter to be accounted for in statistical testing.⁹ With respect to visualization, we plot the nodes in the (rotated) coordinate space formed by the first two eigenvectors, which depicts the network in a way that is inherently related to community structure. The analysis based on the adjacency matrix uses the eigenvectors of that matrix but is otherwise the same as the modularity-based method. Figure 2 schematically depicts the analysis procedure.

Beyond the above techniques exploiting community structure, we also employ a null process simulation to test for homophily conditional upon interaction. This test is effected by randomly permuting the signs of the ties present in the observed network in order to generate a null distribution of a measure of homophily which can then be compared with the observed value. In the null process, the observed set of ties is kept fixed (including multiple ties within dyads) along with the respective numbers of positive and negative ties and then the tie signs are randomly reassigned among ties. The measure of homophily

⁹Appendix D shows the results when correlations are made directly with the top two eigenvectors without rotation.

we use is the assortativity (which is closely related to modularity, Newman (2006)). In an unsigned network, assortativity is equivalent to correlating the values of the variable of interest (e.g., ideology) for connected dyads over all the ties in the network. For a signed network, we use the net assortativity formed by the difference between the separate positive and negative tie assortativities, where each is weighted by their respective fraction of the total ties.¹⁰ Homophily is connoted by an observed assortativity greater than the mean assortativity of the null distribution whereas an observed assortativity less than the null mean connotes heterophily. Sign swapping over fixed tie locations has previously been used to generate null models in studies of structural balance (Leskovec, Huttenlocher, & Kleinberg, 2010; Kirkley, Cantwell, & Newman, 2019).

As an additional robustness check, we perform an exponential random graph model (ERGM) analysis (Desmarais & Cranmer, 2017) (see Appendix E). As the adaptation of ERGMs to signed networks remains nascent, our ERGM analysis treats the cooperative and conflictual networks separately.

Analysis and Results

This section applies the two community structure network analysis methods to our data, followed by the assortativity analysis.

Community Structure Analysis

Figure 3 shows the network for the full time duration of our data as represented in the space formed by the first two eigenvectors of the modularity matrix. Solid lines linking a pair of nodes indicate that the difference between the numbers of cooperative and conflictual events between them is positive and so their relationship is cooperative on net. If this difference is negative, then a dashed line is used indicating a net conflictual relationship. The nodes are shaded with respect to support for Novorossiia, our territorial aspiration variable, with lighter shades indicating higher support. The network has been rotated so that the horizontal axis represents the slice through the network along which the actor coordinates are best correlated with the Novorossiia variable (that the rotation

¹⁰The assortativity for a scalar node variable x for the set of positive ties is defined by $\alpha_+ = \sum_{i,j} B_{ij}^+ x_i x_j / (\sum_{i,j} A_{ij}^+ - k_i^+ k_j^+ x_i x_j / 2m_+)$, where m_+ is the total number of positive ties. The negative tie assortativity α_- is defined analogously. The net assortativity is given by $\alpha = (m_+ \alpha_+ - m_- \alpha_-) / (m_+ + m_-)$.

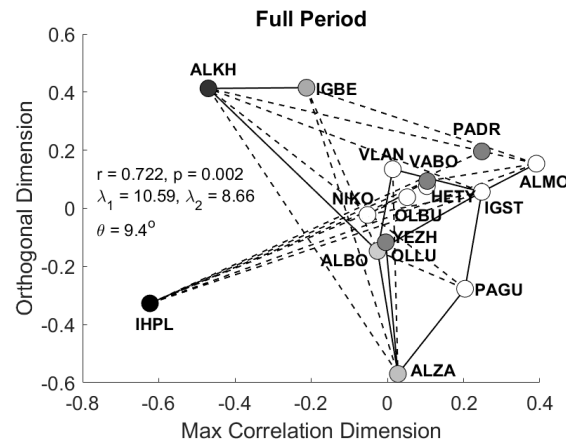


Figure 3: Modularity-based visualization of network for full time period. The network is rotated to maximize correlation with support for Novorossiia. Positive (net cooperative) and negative (net conflictual) ties shown by solid and dashed lines respectively. Nodes are shaded so that lighter shades correspond to more support for the Novorossiia. r is the correlation of the node coordinates along the maximum correlation dimension with the Novorossiia variable (p is the p-value). λ_1 and λ_2 denote the first and second most positive eigenvalues. θ is the angle magnitude by which the eigenvector space must be rotated to achieve r .

angle is small, $\theta = 9.4^\circ$, indicates that the x-coordinates are approximately the same as the components of the first eigenvector).

Visual inspection of the network reveals that the darkest nodes are found on the extreme left whereas the extreme right is comprised primarily of lighter nodes — a pattern suggesting homophily. The right-hand community can be identified as ideologically pro-Novorossiia and three of the four rightmost nodes are ardent Novorossiia supporters, Alexei Mozgovoi (ALMO), Igor Strelkov (IGST), and Pavel Gubarev (PAGU). In contrast, the two leftmost actors, Ihor Plotnytskyi (IHPL) and Alexander Khodakovskiy (ALKH), are the softest on Novorossiia and the left side represents a more pragmatic perspective. This visual indication of homophily is backed up statistically: the correlation r between the Novorossiia variable and the node x-coordinates, shown in the figure and also listed in Table 2, is highly significant ($r = .722$, $p = .002$). This result supports our first hypothesis that state sponsor ambiguity over an ideology component, namely territorial aspiration (Novorossiia), allows that component to shape the network of militant relationships consistent with homophily; ideologically close actors tend to cooperate more and conflict less than do distant actors. The adjacency matrix method for the full time period, further confirms this hypothesis ($r = .672$, $p = .006$, see Appendix C for visualization).

The same network as shown in Figure 3, just shown to be significantly related to territorial aspiration, can also be tested with respect to the conflict frame variable, Russia vs. the West. However, there is no rotation angle for which the correlation is ever significant as Table 2 lists a maximum correlation of $r = .201$ ($p = .47$). This lack of correlation is echoed by the adjacency-based analysis. These null findings support our

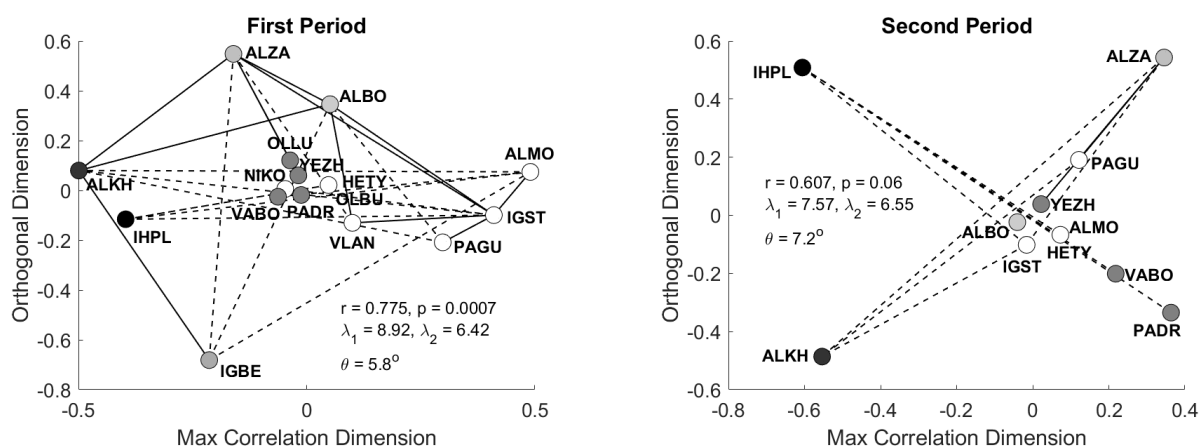


Figure 4: Modularity-based networks for pre and post May 19, 2015 time periods. Left: First time period. Right: Second time period.

second hypothesis, which states that when a sponsor's stance on an ideological component is unambiguous, as the case for Russia regarding the conflict frame, then the network will not display homophily with respect to that component. The conflict frame story is actually more complex and the assortativity analysis below will decisively demonstrate its lack of homophily beyond a simple null finding.

While Russia's conflict frame position was constant, its position on territorial aspiration did change during the span of our data. Accordingly, we divide our data into two time periods, before and after May 19, 2015 when Russia came down against formal secession in order to investigate whether the homophily effect regarding Novorossiia is present in the more limited data period when Russia's position remained ambiguous. We can furthermore investigate whether the territorial aspiration homophily disappears or persists in the second period, after Russia's position became clear. Although naive application of our second hypothesis would predict that homophily would be absent in this period, path-dependent effects may cause relational patterns and ideological positions forged during the initial phase of the conflict to endure for a transient period. Therefore, we do not take the second time period as a definitive test of the second hypothesis.

Figure 4 shows the modularity-based networks for the two time periods with node shadings representing support for Novorossiia. The first period displays homophily — in fact, the correlation is stronger the full period ($r = .775$, $p = .0007$). The adjacency method similarly shows highly significant correlation (Table 2 and Appendix C). The first period analysis therefore thereby confirms that the network homophily aligns with the span of Russian ambiguity.

Homophily is less evident in the second time period. While the most pragmatic actors

Table 2: Network Structure Correlations

	FULL PERIOD			FIRST PERIOD			SECOND PERIOD		
	<i>r</i>	<i>p</i>	θ	<i>r</i>	<i>p</i>	θ	<i>r</i>	<i>p</i>	θ
IDEOLOGY COMPONENTS	$N = 16, m^+ = 65, m^- = 75$			$N = 16, m^+ = 59, m^- = 51$			$N = 11, m^+ = 6, m^- = 24$		
Modularity									
Novorossiia	.722**	.002	9.4	.775***	.0007	5.8	.607	.06	7.2
Russia vs. West	.201	.47	10.5	.149	.60	2.5	.160	.66	42.2
Adjacency									
Novorossiia	.672**	.006	5.8	.732**	.002	11.5	.585	.08	6.5
Russia vs. West	.206	.46	32.1	.236	.40	36.4	.211	.56	62.4

Notes: N is the number of actors in the network. m^+ and m^- are the number of positive and negative ties. N includes only actors with at least one positive or negative tie. $*p < .05$, $**p < .01$, $***p < .001$

(IHPL, ALKH) are still on the far left, the most pro-Novorossiia actors are no longer found on the far right. Indeed, the correlation ($r = .607$, $p = .06$) does not quite attain the standard .05 level of significance. The adjacency-based correlation is also not quite significant ($r = .585$, $p = .08$). Given this near significance, which may reflect the relatively few ties during this period as compared with the first, we do not take the second period Novorossiia as supporting our second hypothesis concerning the lack of homophily when the sponsor is unambiguous. However, our primary test of the second hypothesis involved the conflict frame, for which Russia's position was always clear, and the first and second period correlations for Russia vs. the West remain insignificant (see Table 2) as in the full period analysis.

Finally, we note that the paucity of statements regarding ideal polity and their lack of variation prevented statistical testing against network structure. However, these problems themselves indicate that ideal polity was not salient, which again supports the second hypothesis, given the clear expectation that Russia would not countenance the institution of a radically different form of government in militant-ruled domains.

Assortativity Analysis

In addition to the community structure analysis, summarized in Table 2, as a robustness test we performed the assortativity analysis based on generating a null distribution by randomly exchanging the signs of the ties in the observed network. The results are shown in Table 3. Statistically significant homophily with respect to Novorossiia is observed

Table 3: Assortativity Results

	α	α_{null}	σ_{null}	p
FULL PERIOD				
Novorossiia	.231 (+)**	.010	.069	.001
Russia vs. West	-.306 (-)***	-.030	.078	.0006
FIRST PERIOD				
Novorossiia	.284 (+)***	-.016	.070	<.0001
Russia vs. West	-.199 (-)*	.019	.099	.03
SECOND PERIOD				
Novorossiia	.278 (+)	.087	.119	.10
Russia vs. West	-.658 (-)	-.479	.085	.07

Notes: α is the assortativity of the observed network where + (-) indicates α greater (less) than α_{null} thereby corresponding to homophily (heterophily); α_{null} and σ_{null} are respectively the mean and standard deviation of the assortativity in the null simulation taken over 10,000 runs; the p-value is the (two-tailed) fraction of runs exceeding $|\alpha - \alpha_{null}|$. * $p < .05$, ** $p < .01$, *** $p < .001$

in the full time period as indicated by an observed assortativity $\alpha = .231$ that is more positive than the mean assortativity of the null distribution, $\alpha_{null} = .01$, and is over three standard deviations away. The first period shows even higher assortativity. These results match well with the initial community structure findings, providing strong confirmation of our first hypothesis: state sponsor ambiguity produces network homophily with respect to the ambiguous component.

With respect to the conflict frame of Russia vs. the West, the assortativity analysis fails to find homophily, in accordance with our hypothesis. In fact, it shows highly significant heterophily: the observed assortativity in the full period, $\alpha = -.306$, is 3.5 standard deviations less than the null assortativity, $\alpha_{null} = -.03$. We can also look for this heterophily with the community structure analysis by using the last two eigenvectors, i.e., those having the two most negative eigenvalues, instead of the first two as for homophily. The correlation between Russia vs. the West and the second most negative eigenvector of the modularity matrix is highly significant ($r = .759$, $p = .0007$, see Appendix F for visualization) and so confirms the assortativity result.

This finding of heterophily goes beyond the mere absence of homophily as predicted by the second hypothesis. Its implication — that having more similar conflict frame positions tends to produce less cooperation and greater conflict — is strange and surprising, even more so given the relative lack of variation along Russia vs. the West as compared

with Novorossiia (see Data and Sources/Ideology). We offer a candidate explanation presently, but regardless of the ultimate explanation, the empirical finding that these two ideological components can have dramatically different effects on militant factional behavior (homophily for territorial aspiration and heterophily for conflict frame) points to the importance of disaggregating ideology into distinct dimensions rather than treating it as a single variable.

Our speculative explanation for the heterophily finding relies on the distinction made in our theory regarding the actionable nature of an ideological component for which the sponsor's position is ambiguous as opposed to the rhetorical nature of a component for which the sponsor's position is clear. In the Donbass case, the territorial aspiration component is actionable: a militant leader's position regarding Novorossiia has very concrete implications as to ends and means, such as whether to push for further territorial gains or to support the Minsk ceasefire agreements. Consequently, having similar Novorossiia positions served as a spur to joint action, such cooperation being the core impetus underlying ideological homophily in general. In contrast, the conflict frame was primarily rhetorical with little impact on militant actions. However, variation in militant leader positions along the Russia vs. the West frame could indicate differences in their constituencies among rank-and-file fighters, the populace, or the Russian government itself. Leaders with similar positions would therefore have similar constituencies, which could serve as a source of competition and conflict, thereby promoting heterophily.¹¹

The ERGM analysis supports our hypotheses despite the limited statistical power associated with treating the cooperative and conflictual networks separately. While the cooperative network yields no statistical significance, the conflictual network results aligns with the analysis above, exhibiting significant homophily with respect to Novorossiia and heterophily with respect to Russia vs. the West in both the full and first time periods (see Appendix E).

Conclusion

Understanding the dynamics of cooperation and conflict between groups in an insurgent movement is a critical task for the study of modern conflict. The last decade has seen an

¹¹The homophily effect of Novorossiia is a stronger driver of network structure than the heterophily of Russia vs. the West as its associated eigenvalue is of larger magnitude.

impressive growth in scholarship on this topic. We contribute by presenting a theoretical explanation of conflict and cooperation dynamics that emphasizes the role of external sponsorship and ideology. We argue that ideological homophily is impacted by the number and ideological stances of external sponsors. Central to this account, especially in the single sponsor case that we highlight, is the consideration of insurgent ideology in terms of three distinct dimensions — conflict frame, ideal polity, and territorial aspiration — rather than treating it as an aggregate variable. Employing this framework, we have argued that in the presence of a single external sponsor, ideological components along which the sponsor is ambiguous will shape militant networks.

We tested the theory in the case of Ukraine, where an insurgent conflict starting in early 2014 pitted militants backed by Russia against the Ukrainian state. This is a tough case for the claim that ideology matters in that one would expect little room for ideological variation in an insurgency backed by one powerful external sponsor. Contrary to these expectations, we showed that network homophily is indeed correlated with actors' positions on territorial aspiration (Novorossiia), the ideological component about which Russia initially sent ambiguous signals. Statistically significant homophily was found using all of our analysis methods — modularity and adjacency-based community structure analysis, assortativity, and ERGMs — for the full time period of our data as well as its restriction to just the time before Russia finally shed its ambiguity, coming down against formal secession from Ukraine.

We also showed that conflict frame, which according to our theory should not generate homophily in this case because Russia always exhibited a clear stance, indeed did not do so. In fact, going beyond the mere absence of a relationship with network structure, the conflict frame exhibited heterophily, a counterintuitive behavior for ideology, which we speculated stemmed from the rhetorical nature of sponsor-unambiguous ideological components. Regardless of the causal mechanism and whether such heterophily commonly occurs, its observation in the Ukraine case further speaks to the value of disaggregating ideology in that different components may actually behave in opposite ways.

Methodologically, our results show the promise of signed network analysis for the study of militant dynamics. Signed networks allow for an integrated analysis of cooperation and conflict, behaviors which are deeply intertwined. The community structure approach, whether adjacency or modularity-based, allows for the statistical analysis of the correlates

of network homophily and for meaningful visual representation of that structure. The assortativity analysis, which generates a null distribution by randomly exchanging signs among existing ties, tests homophily conditional upon actor interaction and therefore can directly show that homophily results from actor preferences rather than a lack of diversity in potential partners, unlike unsigned networks.

As our paper has been limited to a single, in-depth case study, further testing of our theory is warranted for both single and multiple-state sponsor cases. For additional cases, researchers might consider the proxy insurgent movements sponsored by the United States and the Soviet Union during the Cold War for evidence of the effects of state sponsor ambiguity over key militant ideological dimensions. One such case would be U.S. support for the Contra rebels that fought the Sandinista regime in Nicaragua from 1980-1989. After the fall of the Somoza government and establishment of the Sandinista government, ex-military exiles received logistic, financial, and material support from the U.S. government. Although other minor sponsors were present, they were essentially surrogates of the U.S., which effectively was the single state sponsor. In fact, U.S. support was so crucial to the Contras that, when it was removed due to scandal and changing U.S. political priorities, the Contras could not continue the fight and a negotiated settlement followed quickly (DeRouen & Heo, 2007, p. 553).

The Contras were not a monolithic bloc and were comprised of distinct groups such as military officers and capitalist economic elites affiliated with the deposed authoritarian regime, socialist-leaning former rebels who soured on the Sandinistas after they gained power; and indigenous populations agitating for greater political rights and power. Aside from being anti-Sandinista, the various groups fighting the government were at times either vague or at odds with each other on why they were motivated to fight or what a post-Sandinista Nicaragua would look like. These differences were wholly unimportant to the Reagan Administration, which was interested in undermining the Sandinista government and stamping out leftist politics in Central America, consistent with its Cold War strategy (Ronfeldt & Jenkins, 1989, p. 26). To that end, the U.S. had a clearly defined out-group conflict frame, which was a point of agreement among all parts of the Contra movement, but as to in-group or ideal polity, the U.S. government in the 1980s was often either disinterested or fragmented (Ronfeldt & Jenkins, 1989, pp. 15, 17-18), which, according to our theory, would allow the groups to exhibit diversity along these elements of ideology.

Further study could investigate whether the patterns of cooperation and conflict between the various Contra factions were predicted by variation on the ideological components of ideology about which the U.S. was ambiguous.

Although further empirical support for the generality of our theory must await such additional comparative analyses, our investigation of Eastern Ukraine has extended the scope of conflicts for which ideology should be treated as relevant to shaping militant conflict and cooperation, which previously had been observed only in the Syrian civil war networks (Gade, Hafez, & Gabbay, 2019; Gade, Gabbay, et al., 2019). In addition, we have put forth a broad theory of the interaction between external sponsorship, ideology, and militant network structure that we believe will spur future research. Our findings suggest that future researchers interested in the effects of ideology and state sponsorship should focus their attentions on ideological dimensions where either the single external state sponsor is ambiguous or where multiple sponsors disagree on ideological components.

References

- Ambrosio, T. (2016). The rhetoric of irredentism: The Russian Federation's perception management campaign and the annexation of crimea. *Small Wars & Insurgencies*, 27(3), 467–490.
- Bakke, K. M., Cunningham, K. G., & Seymour, L. J. M. (2012). A plague of initials: Fragmentation, cohesion, and infighting in civil wars. *Perspectives on Politics*, 10(2), 265–283.
- Bapat, N. A. & Bond, K. D. (2012). Alliances between militant groups. *British Journal of Political Science*, 42(04), 793–824.
- Bowen, A. S. (2019). Coercive diplomacy and the Donbas: Explaining Russian strategy in Eastern Ukraine. *Journal of Strategic Studies*, 42(3-4), 312–343.
- Brandt, P. T., Mason, T. D., Gurses, M., Petrovsky, N., & Radin, D. (2008). When and How the Fighting Stops: Explaining the Duration and Outcome of Civil Wars. *Defence and Peace Economics*, 19(6), 415–434.
- Byman, D., Chalk, P., Hoffman, B., Rosenau, W., & Brannan, D. (2001). *Trends in Outside Support for Insurgent Movements*. RAND. Santa Monica, CA.

- Cartwright, D. & Harary, F. (1956). Structural balance: A generalization of heider's theory. *Psychological Review*, 63(5), 277–293.
- Christia, F. (2012). *Alliance Formation in Civil Wars*. Cambridge: Cambridge University Press.
- Costalli, S. & Ruggeri, A. (2015). Indignation, Ideologies, and Armed Mobilization: Civil War in Italy, 1943–45. *International Security*, 40(2), 119–157.
- Cunningham, D. E. (2010). Blocking resolution: How external states can prolong civil wars. *Journal of Peace Research*, 47(2), 115–127.
- DeRouen, K. R. & Heo, U. (2007). *Civil wars of the world: Major conflicts since World War II*. Santa Barbara, Calif.: ABC-CLIO.
- Desmarais, B. A. & Cranmer, S. J. (2017). Statistical inference in political networks research. In J. N. Victor, A. H. Montgomery, & M. Lubell (Eds.), *The Oxford Handbook of Political Networks*. Oxford University Press.
- Dorff, C., Gallop, M., & Minhas, S. (2020). Networks of violence: Predicting conflict in nigeria. *The Journal of Politics*, 82(2), 476–493.
- East Ukraine separatists seek union with Russia. (2014). *BBC*.
- Fjelde, H. & Nilsson, D. (2018). The rise of rebel contenders: Barriers to entry and fragmentation in civil wars. *Journal of Peace Research*, 55(5), 551–565.
- Gade, E. K., Gabbay, M., Hafez, M. M., & Kelly, Z. (2019). Networks of Cooperation: Rebel Alliances in Fragmented Civil Wars. *Journal of Conflict Resolution*, 002200271982623.
- Gade, E. K., Hafez, M. M., & Gabbay, M. (2019). Fratricide in rebel movements: A network analysis of Syrian militant infighting. *Journal of Peace Research*, 002234331880694.
- Gall, C. (2018). Ukraine Town Bears Scars of Russian Offensive That Turned Tide in Conflict. *The New York Times*.
- Gutiérrez Sanín, F. & Wood, E. J. (2014). Ideology in civil war: Instrumental adoption and beyond. *Journal of Peace Research*, 51(2), 213–226.

- Hirose, K., Imai, K., & Lyall, J. (2017). Can civilian attitudes predict insurgent violence? Ideology and insurgent tactical choice in civil war. *Journal of Peace Research*, 54(1), 47–63.
- Hutchings, S. & Szostek, J. (2016). Dominant Narratives in Russian Political and Media Discourse during the Ukraine Crisis. In A. Pikulicka-Wilczewska & R. Sakwa (Eds.), *Ukraine and Russia: People, politics, propaganda and perspectives*. OCLC: 965380133.
- International Crisis Group. (2014). *Eastern Ukraine : A Dangerous Winter* (tech. rep. No. December).
- International Crisis Group. (2019). Rebels without a Cause: Russia's Proxies in Eastern Ukraine. (254), 30.
- Karlén, N. (2017). The Legacy of Foreign Patrons: External State Support and Conflict Recurrence. *Journal of Peace Research*, 54(4), 499–512.
- Kirkley, A., Cantwell, G. T., & Newman, M. E. J. (2019). Balance in signed networks. *Physical Review E*, 99(1), 012320.
- Kramer, A. E. & Gordon, M. R. (2014). Russian troops launch stealth invasion, Ukraine says. *Boston Globe*.
- Krause, P. (2014). The Structure of Success: How the Internal Distribution of Power Drives Armed Group Behavior and National Movement Effectiveness. *International Security*, 38(3), 72–116.
- Krause, P. (2017). *Rebel power: Why national movements compete, fight, and win*. Ithaca, NY: Cornell University Press.
- Kudelia, S. (2019). How they joined? Militants and informers in the armed conflict in Donbas. *Small Wars & Insurgencies*, 30(2), 279–306.
- Laruelle, M. (2016). The three colors of Novorossiia, or the Russian nationalist mythmaking of the Ukrainian crisis. *Post-Soviet Affairs*, 32(1), 55–74.
- Leader Maynard, J. (2019). Ideology and armed conflict. *Journal of Peace Research*, 56(5), 635–649.
- Leskovec, J., Huttenlocher, D., & Kleinberg, J. (2010). Signed networks in social media. In *Proceedings of the sigchi conference on human factors in computing*

- systems* (pp. 1361–1370). CHI '10. Atlanta, Georgia, USA: Association for Computing Machinery.
- Lidow, N. H. (2016). *Violent order: Understanding rebel governance through Liberia's civil war*. New York, NY: Cambridge University Press.
- MacFarquhar, N. (2014). Russia Keeps Its Distance After Ukraine Secession Referendums. *The New York Times*.
- Maoz, Z. (2011). *Networks of Nations*. Cambridge: Cambridge University Press.
- Maoz, Z., Terris, L. G., Kuperman, R. D., & Talmud, I. (2007). What is the enemy of my enemy? Causes and consequences of imbalanced international relations, 1816–2001. *The Journal of Politics*, 69(1), 100–115.
- Matveeva, A. (2016). No Moscow Stooges: Identity Polarization and Guerrilla Movements in Donbass. *Journal of Southeast European and Black Sea*, 16(1), 25–50.
- Metternich, N. W., Dorff, C., Gallop, M., Weschle, S., & Ward, M. D. (2013). Antigovernment networks in civil conflicts: How network structures affect conflictual behavior. *American Journal of Political Science*, 57(4), 892–911.
- Minhas, S., Hoff, P. D., & Ward, M. D. (2016). A New Approach to Analyzing Co-evolving Longitudinal Networks in International Relations. *Journal of Peace Research*, 53(3), 491–505.
- Newman, M. E. J. (2006). Finding Community Structure in Networks using the Eigenvectors of Matrices. *Physical Review E*, 74(3), 36104.
- O'Loughlin, J., Toal, G., & Kolosov, V. (2017). The rise and fall of “Novorossiya”: Examining support for a separatist geopolitical imaginary in southeast Ukraine. *Post-Soviet Affairs*, 33(2), 124–144.
- Oppenheim, B., Steele, A., Vargas, J. F., & Weintraub, M. (2015). True Believers, Deserters, and Traitors: Who Leaves Insurgent Groups and Why. *The Journal of Conflict Resolution*, 59(5), 794–823.
- Popovic, M. (2018). Inter-rebel alliances in the shadow of foreign sponsors. *International Interactions*, 44(4), 749–776.

- Regan, P. M. (2002). Third-Party Intervention and the Duration of Intrastate Conflicts. *Journal of Conflict Resolution*, 46(1), 55–73.
- Ronfeldt, D. & Jenkins, B. M. (1989). The Nicaraguan Resistance and U.S. Policy: Report on a May 1987 Conference. Library Catalog: www.rand.org Publisher: RAND Corporation.
- Salehyan, I., Gleditsch, K. S., & Cunningham, D. E. (2011). Explaining external support for insurgent groups. *International Organization*, 65(4), 709–744.
- San Akca, B. (2016). *States in Disguise: Causes of State Support for Rebel Groups*. New York: Oxford University Press.
- Sawyer, K., Cunningham, K. G., & Reed, W. (2017). The Role of External Support in Civil War Termination. *Journal of Conflict Resolution*, 61(6), 1174–1202.
- Shesterinina, A. (2016). Collective threat framing and mobilization in civil war. *American Political Science Review*, 110(3), 411–427.
- Sinno, A. H. (2008). *Organizations at War in Afghanistan and Beyond*. Cornell University Press.
- Staniland, P. (2014). *Networks of Rebellion: Explaining Insurgent Cohesion and Collapse*. Ithaca, NY: Cornell University Press.
- Tamm, H. (2016). Rebel Leaders, Internal Rivals, and External Resources: How State Sponsors Affect Insurgent Cohesion. *International Studies Quarterly*, 60(4), 599–610.
- Traag, V. & Bruggeman, J. (2009). Community Detection in Networks with Positive and Negative Links. *Physical Review E*, 80(3), 36115.
- Traag, V., Doreian, P., & Mrvar, A. (2020). Partitioning signed networks. In P. Doreian, V. Batagelj, & A. Ferligoj (Eds.), *Advances in Network Clustering and Blockmodeling* (pp. 225–249). Hoboken, NJ: John Wiley Sons.
- Ugarriza, J. E. & Craig, M. J. (2013). The Relevance of Ideology to Contemporary Armed Conflicts: A Quantitative Analysis of Former Combatants in Colombia. *Journal of Conflict Resolution*, 57(3), 445–477.
- Williams, C. J. (2014). Russia Recognizes Ukraine Separatists, Provoking New Sanctions Threat.

- Zaken, M. v. A. (2018). The Netherlands holds Russia responsible for the MH17 tragedy - MH17 incident - Government.nl. onderwerp.
- Zhukov, Y. M. (2016). Trading hard hats for combat helmets: The economics of rebellion in eastern Ukraine. *Journal of Comparative Economics*, 44(1), 1–15.

Appendix

A Actor list

Our initial actor list included all of the groups and individuals noted by Zhukov (2016) and Matveeva (2016). We then added all groups on the Wikipedia pages for the Donbass conflict.¹² As we began manually coding documents, we developed criteria for adding new actors. First, to be added to the actor list, an actor had to appear in more than one document. Second, actors who were solely political, with no apparent connection to militant activity, were excluded. A connection to militant activity could include financing a group or being rumored to be closely connected to a militant leader. We collected data on 28 unique insurgent leaders out of 258 groups and individuals that we identified as related to the conflict, not all of whom were insurgent leaders.

We assigned each actor in the list a unique four letter code. This code is generally the first two letters of an individual's first name plus the first two letters of their last name (unless that combination was already taken or unless a particular callsign or nickname was entered before we knew the actual actor name). For example, Igor Strelkov is represented in the network graphics as IGST, but Botsman is BOTS.

One challenge we faced was in different transliterations of names (often a Russian versus Ukrainian variant) and the widespread practice of militants using callsigns or nicknames. Igor Strelkov's real name, for example, is Igor Girkin. Strelkov is a nickname that comes from the Russian word for "shot" or "shooter." We linked his name and nickname to a range of alternative spellings noted in the documents, including: Igor Girkin, Ihor Strelkov, Ihor Girkin, Strelok, Ihor Hirkin, Ihor Hyrkin, and Strelkov-Hyrkin. We noted alternative spellings and callsigns and ensured that they were assigned to only one actor.

In order to reduce the number of actors in the network, where possible we consolidated groups to individuals. This reflected our observation that more often the

¹²e.g. https://en.wikipedia.org/wiki/War_in_Donbass, https://en.wikipedia.org/wiki/Separatist_forces_of_the_war_in_Donbass, and [https://en.wikipedia.org/wiki/Ukrainian_volunteer_battalions_\(since_2014\)](https://en.wikipedia.org/wiki/Ukrainian_volunteer_battalions_(since_2014))

articles described individuals, as opposed to militant groups. Thus, for example, all references to the Vostok battalion were ultimately assigned to Alexander Khodakovsky, the group's leader. When the leadership was unclear, or when it shifted over time, we did not consolidate a group to an individual. In the events database, after actor consolidation, we have 28 unique actors.¹³ Table A1 shows all of the included actors, their network codes, and which specific side of the conflict the actor is on (LNR or DNR).

Table A1: Actors in the Network

Network Code	Name	Specific Side
ALBO	Aleksander Borodai	Donetsk
BATM	Alexander Bednov	Luhansk
ALKH	Alexander Khodakovsky	Donetsk
ALZA	Alexander Zakharchenko	
ALMO	Alexei Mozgovoi	Luhansk
ARPA	Arseny Pavlov	Donetsk
BOTS	Botsman	Donetsk
BURY	Buryat	
HETY	Hennadiy Tsypkalov	Luhansk
IGBE	Igor Bezler	Donetsk
IGST	Igor Strelkov	Donetsk
IHPL	Ihor Plotnytskyi	Luhansk
MITO	Mihail Tolstyh	Donetsk
NIKO	Nikolai Kozitsyn	Luhansk
OLBU	Oleh Buhrov	Luhansk
OLLU	Oleksandr Lukyanchenko	Donetsk
PAGU	Pavel Gubarev	Donetsk
PADR	Pavlo Dryomov	Luhansk
SEPE	Sergei Petrovskiy	
SEZD	Sergey Zdrylyuk	Donetsk
SELI	Serhiy Lytvyn	Luhansk
VABO	Valerii Bolotov	Luhansk
VLAN	Vladimir Antyufeyev	Donetsk
VLKO	Vladimir Kononov	Donetsk
VYPO	Vyacheslav Ponomaryov	Donetsk
YEZH	Yevgenyi Zhilin	Donetsk
YEIS	Yevhen Ishchenko	Luhansk
YEKL	Yevhen Klep	Donetsk

¹³These actors came from a larger list of 258 groups and individuals that we identified as being related to the conflict, but not all of whom are insurgent leaders.

B Event Coding

An event in our database follows the structure of “someone (the “from” actor) doing something (the event action) to someone else (the “to” actor).” We only coded events where both the “from” and “to” actors were included in our list of actors to track. If an event description included more than one actor (either “from” or “to”), we coded each from-to pair as a separate event. For each event, we also noted the time of the event (to the day, if possible, or to the month or year if no more specific date was given) and the published date of the source for the event.

Each event was given three event type codes, at different levels of specificity. First is a general event type, either **conflict** or **cooperation**. A particular report of an event was coded for as many different dyadic interactions as the event contained. So, for example, a report of two insurgents cooperating to attack another insurgent group would contain three events; one cooperative and two conflictual.¹⁴ We further classify an event subtype for each act of cooperation or conflict as either rhetorical, violent, material/military, or relational. **Rhetorical** events are those in which one actor is saying something about another actor. This can be praise (in which case the event is cooperation) or criticism (in which the event is conflict) or any other relevant type of statement. **Violent** events involved actual use of force of one actor against another actor (conflict) or actual use of force in which one actor is fighting alongside another actor (cooperation). **Material/military** events are those that involve one actor engaging in some action that is greater than mere rhetorical interaction with another actor but that does not fall under what has been described as a violent event. Promoting, financing, demoting, and protecting all count as material/military events. **Relational** events include any references to a relationship (cooperative or conflictual) between two actors. In this field we also tagged whether or not the event was being described by a “3rd party” source, meaning either that the source was describing something the author did not witness firsthand or that the source was not trustworthy (e.g., one commander claiming that two other commanders hated each other would be coded as a 3rd party rhetorical

¹⁴Cooperative (1): insurgent A cooperates with insurgent B to attack insurgent C. Conflictual (2): insurgent A attacks insurgent C; insurgent B attacks insurgent C.

conflict between the two commanders supposedly at odds). Tables B1 and B2 show the number of events in each subtype.

Table B1: Cooperation Subtypes

Event Subtype	Count
material/military	37
relational	22
rhetorical	15
violent	12

Table B2: Conflict Subtypes

Event Subtype	Count
material/military	28
relational	2
rhetorical	38
violent	17

Since we were coding media reports, information that appeared in one story frequently appeared in another. To deal with the problem of double- or triple-coding the same event, we manually reviewed the events database for duplicate events and removed them. In total, coding the 136 documents yielded 171 total events involving relevant actors, of which 86 were instances of cooperation and 85 were conflict.

C Adjacency-Based Visualizations

The figures show the networks obtained by rotating the eigenvectors of the adjacency matrix and parallel the modularity-based versions in the main text.

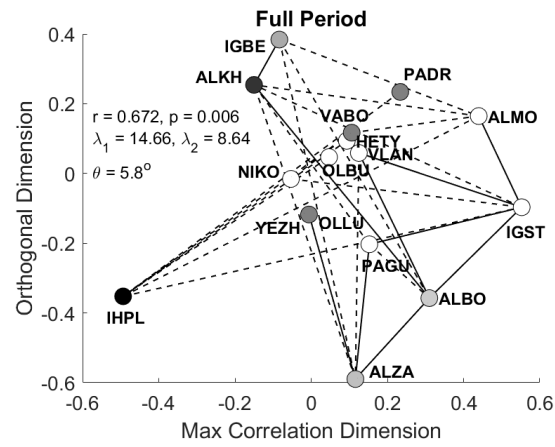


Figure 5: Adjacency-based visualization of network for full time period.

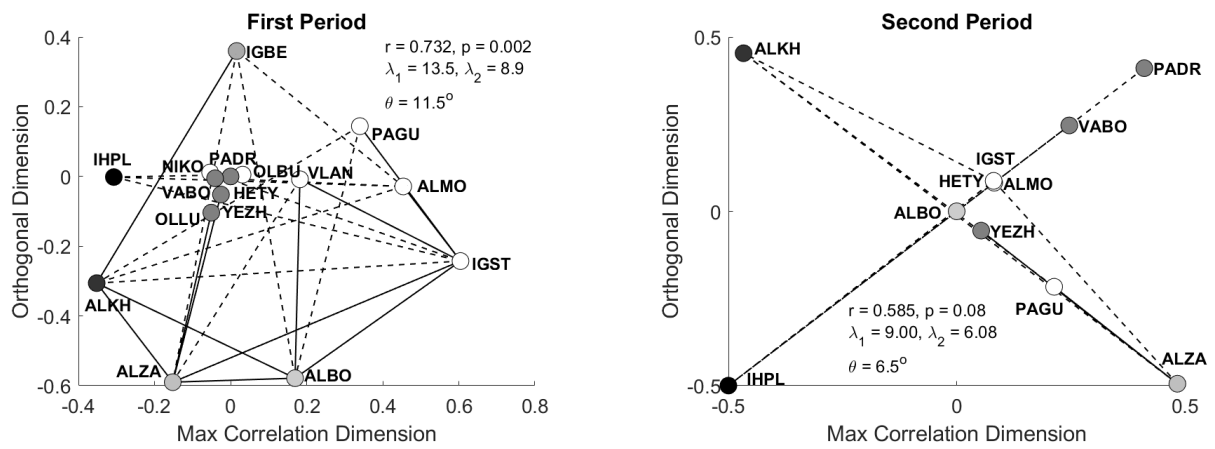


Figure 6: Network structure for pre and post May 19, 2015 time periods using adjacency matrix eigenvectors.

D Analysis Results without Network Rotation

Table B3: Network Eigenvector Correlations without Rotation

	FULL PERIOD			FIRST PERIOD			SECOND PERIOD		
	<i>r</i>	<i>p</i>	<i>EV</i>	<i>r</i>	<i>p</i>	<i>EV</i>	<i>r</i>	<i>p</i>	<i>EV</i>
IDEOLOGY COMPONENTS	<i>N</i> = 16, <i>m</i> ⁺ = 65, <i>m</i> ⁻ = 75			<i>N</i> = 16, <i>m</i> ⁺ = 59, <i>m</i> ⁻ = 51			<i>N</i> = 11, <i>m</i> ⁺ = 6, <i>m</i> ⁻ = 24		
Modularity									
Novorossiya	.705**	.002	1	.662**	.005	1	.439	.18	2
Russia vs. West	.143	.60	1	.145	.59	1	.146	.67	2
Adjacency									
Novorossiya	.569*	.02	1	.648**	.007	2	.41	.21	2
Russia vs. West	.191	.48	2	.188	.49	1	.157	.65	2

Notes: *EV* lists which of the first two eigenvectors correlates best with the ideology component. Other quantities as in main text. **p* < .05, ***p* < .01, ****p* < .001

E ERGM

E.1 Full period analysis

	Model 1
edges	−1.05* (0.50)
absdiff.pro_ru	−2.27*** (0.69)
absdiff.pro_novo	1.16** (0.44)
AIC	87.98
BIC	95.51
Log Likelihood	−40.99

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table B4: Full period conflict network

	Model 1
edges	−0.09 (0.56)
absdiff.pro_ru	−1.23 (0.76)
absdiff.pro_novo	−0.83 (0.62)
AIC	73.99
BIC	80.56
Log Likelihood	−34.00

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table B5: Full period cooperation networks

E.2 Period 1 analysis

	Model 1
edges	−1.15* (0.51)
absdiff.pro_ru	−2.10** (0.68)
absdiff.pro_novo	1.12* (0.44)
AIC	88.07
BIC	95.60
Log Likelihood	−41.03

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table B6: Period 1 conflict networks

	Model 1
edges	−0.28 (0.56)
absdiff.pro_ru	−1.03 (0.76)
absdiff.pro_novo	−0.78 (0.62)
AIC	72.83
BIC	79.39
Log Likelihood	−33.41

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table B7: Period 2 cooperation networks

E.3 Period 2 analysis

	Model 1
edges	−2.22* (1.02)
absdiff.pro_ru	−2.68* (1.21)
absdiff.pro_novo	2.08* (0.84)
AIC	38.88
BIC	44.30
Log Likelihood	−16.44

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table B8: Period 2 conflict networks

	Model 1
edges	−0.28 (1.45)
absdiff.pro_ru	0.27 (2.09)
absdiff.pro_novo	−0.41 (2.13)
AIC	19.40
BIC	20.31
Log Likelihood	−6.70

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table B9: Period 2 cooperation networks

F Heterophily in Russia vs. the West

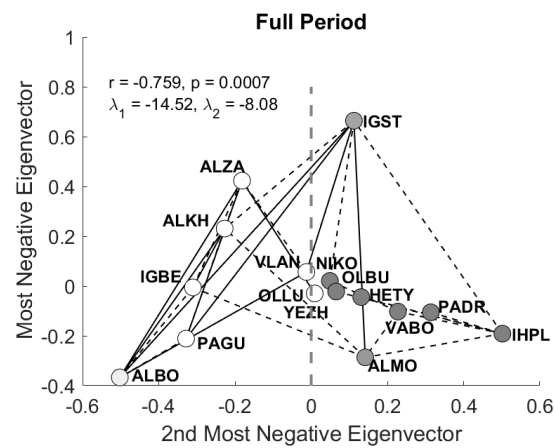


Figure 7: Visualization of network for full time period showing heterophily for the conflict frame. The visualization is based on the two lowest eigenvectors of the modularity matrix, i.e., those with the first and second most negative eigenvalues, λ_{1MN} and λ_{2MN} respectively. Lighter nodes indicate greater support for the Russia vs. the West frame.