Supplementary information for "Network polarization: The study of political attitudes and social ties as dynamic multilevel networks"

Kieran Mepham[1*], András Vörös[2], and Christoph Stadtfeld[1]

[1] Social Networks Lab, ETH Zurich, Switzerland
[2] School of Social Policy, University of Birmingham, UK
[*]Corresponding author. Email: kieranmepham@gmail.com

Summary

The Supplementary Information relate to the article "Network polarization: The study of political attitudes and social ties as dynamic multilevel networks" (published in *Network Science*, 2025). It first presents mathematical and computational details on the calculation of the "Network Polarization" metric that the article introduces (appendix A), then shows details on the measures and descriptives of the political items battery (appendices B and C), and then adds full models, network descriptives, goodness-of-fit measures and robustness analyses that relate to the Stoachstic Actor-oriented Models presented in the main paper (appendices D–L).

A Network polarization calculation

Using the networks of bipartite attitudes and people, and people amongst one another, we define the metrics capturing network polarization. Figure 2 in the main text shows stylized forms of a non-polarized network, and a polarized network under our metrics and networks that are only partially polarized. These metrics are constructed from specific network structures - the census of mixed triads and the census of mixed, complete four-cycles of people, friendships, political statements and attitudes. Censuses of network structures have a relatively long history in the social networks literature both in descriptive and in inferential analyses (for seminal publications, see Davis, 1970; Holland and Leinhardt, 1977). We thus take a census of the possible configurations of these complete four-cycles and triads, and examine whether their prevalence relative to one another would suggest polarization in our data.

To do this, we also define an expectation. In the ideological aspect of polarization, we define this expectation from a sample of synthetic opinion networks with the same degree distribution (i.e. all nodes have the same degree) as the observation. To find these networks, we perform a rewiring procedure. Step by step, this occurs as follows:

1. Select a random individual *a1* (weighted by number of ties)
2. Select a random attitudinal object *b1* it is connected to
3. Check the valence of the tie (positive or negative)
4. Select a random connected individual-attitude pair *a2b2* of which neither is connected to the individual-attitude pair from steps 1 and 2, with the same valence as tie *a1b1*
5. Swap the ties so that one goes from a1 to b2, and the other goes from a2 to b1
6. Repeat for 3 times the number of ties observed in both the positive and negative network

For the relational aspect of polarization we take a simpler approach, and simultaneously permute the rows and columns of the friendship adjacency matrix, while holding the opinion network constant. This implies an identical structure of the friendship network, but a changed relationship between the opinion and friendship network.

We generate 1000 synthetic opinion networks, and 1000 synthetic friendship networks using this procedure. All code to generate these is available on the github repository: [omitted for review, available on request] Functions 'rewire3' and 'permuteNet' generate the networks following the procedures above.

To capture the ideological facet of polarization, defined in Formula 1 we count four-cycles of opinions, specifically we take a census of complete tetrads containing two individuals and two topics. Only opinion ties are considered, while friendship ties are ignored in these configurations. For both the observed and synthetic opinion networks, we calculate statistics given below.

i and j are individuals where $i \neq j$, k and m where $k \neq m$ are political issues, and w_+ and w_- indicate positive or negative opinion ties between individuals and political issues.

$$s_{\text{agree}} = \sum w_{ik_+} w_{jk_+} w_{im_+} w_{jm_+} + \sum w_{ik_-} w_{jk_-} w_{im_-} w_{jm_-} + \sum w_{ik_+} w_{jk_+} w_{im_-} w_{jm_-} \tag{1a}$$

defines the count of tetrads where individuals completely agree, while

$$s_{\text{disagree}} = \sum w_{ik_+} w_{jk_-} w_{im_+} w_{jm_-} + \sum w_{ik_+} w_{jk_-} w_{im_-} w_{jm_+} \tag{1b}$$

defines the count of tetrads where individuals completely disagree. Finally,

$$s_{\text{inconsistent}} = \sum w_{ik_+} w_{jk_+} w_{im_+} w_{jm_-} + \sum w_{ik_-} w_{jk_-} w_{im_-} w_{jm_+} \tag{1c}$$

defines the count of tetrads where individuals agree on one issue but disagree on another.

Each term is then normalized, by dividing by the expectation. This is the mean of each statistic across the 1000 synthetic networks described above. For example, s_{agree} is divided by $E(s_{agree})$.

After normalization, we define each metric as follows:

$$s_{\text{ideological attraction}} = \frac{s_{\text{agree}}}{s_{\text{agree}} + s_{\text{inconsistent}}} \tag{1d}$$

$$s_{\text{ideological repulsion}} = \frac{s_{\text{disagree}}}{s_{\text{disagree}} + s_{\text{inconsistent}}} \tag{1e}$$

$$s_{\text{ideological polarization}} = \frac{s_{\text{ideological attraction}} + s_{\text{ideological repulsion}}}{2} \tag{1f}$$

To capture the social facet of polarization, represented in Formula 2, we start with the census of triads containing two people, one attitudinal object, and two opinion ties. We then calculate the following statistics for both the observed

network of friendships and opinions, and the combined observed opinions with 1000 permuted friendship networks.
x_{ij} represents the binary (0 for no tie, 1 for a tie) value of directed friendships between individuals ij, where $i \neq j$. w_{k_+} represents positive attitudinal ties to issue k, and w_{k_-} represents negative attitudinal ties to issue k.

$$s_{\text{friends agree}} = \sum x_{ij} w_{ik_+} w_{jk_+} + \sum x_{ij} w_{ik_-} w_{jk_-} \tag{2a}$$

defines the count of triads of friends with valenced agreement on a topic,

$$s_{\text{friends disagree}} = \sum x_{ij} w_{ik_-} w_{jk_+} + \sum x_{ij} w_{ik_+} w_{jk_-} \tag{2b}$$

defines the count of triads of friends with valenced disagreement on a topic,

$$s_{\text{nonfriends disagree}} = \sum -1 * (x_{ij} - 1) w_{ik_-} w_{jk_+} + \sum -1 * (x_{ij} - 1) w_{ik_+} w_{jk_-} \tag{2c}$$

defines the count of triads of unconnected individuals with valenced disagreement on a topic

$$s_{\text{nonfriends agree}} = \sum -1 * (x_{ij} - 1) w_{ik_+} w_{jk_+} + \sum -1 * (x_{ij} - 1) w_{ik_-} w_{jk_-} \tag{2d}$$

defines the count of triads of unconnected individuals with valenced agreement on a topic.

We normalize all statistics by the expectation, as above. Then, we calculate the relational polarization indices as follows:

$$s_{\text{social attraction}} = \frac{s_{\text{friends agree}}}{s_{\text{friends agree}} + s_{\text{nonfriends agree}}} \tag{2e}$$

$$s_{\text{social repulsion}} = \frac{s_{\text{nonfriends disagree}}}{s_{\text{nonfriends disagree}} + s_{\text{friends disagree}}} \tag{2f}$$

$$s_{\text{relational polarization}} = \frac{s_{\text{social attraction}} + s_{\text{social repulsion}}}{2} \tag{2g}$$

Both metrics theoretically range from zero to one. It has not been explored yet to what extent the structure of the observed networks may affect the empirically possible range of the metrics.

B Political statement battery

The presented political items battery was informed by different sources as outlined in the main article. Details on the research design and data collection can be found in Vörös et al. (2021).

1. There should be an annual upper limit to the uptake of new asylum seekers.
2. Switzerland should be able to set quotas for the immigration of foreign workers.
3. Operators of internet sites should be legally mandated to remove Fake News that they are made aware of.
4. Children must be vaccinated against contagious illnesses.
5. All banks should be nationalized.
6. There should be a quota for the number of women on supervisory boards of listed companies.
7. High wealth should be taxed.
8. Arms exports from Switzerland should be prohibited, without exception.
9. The regulated sale of cannabis should be generally permitted.
10. Bank customer secrecy privileges in Switzerland should be loosened to combat tax evasion.
11. The state should be allowed to invade the privacy of citizens more forcefully to protect against terrorism.
12. Video surveillance in public spaces should be expanded.
13. Radio and TV license fees should be abolished and replaced by commerical financing.
14. The expansion of renewable energy sources should receive permanent financial support from the state.
15. Electromobility should be promoted more than conventional mobility technologies.
16. More bike paths should be built on public roads in Switzerland.
17. Organic agriculture should be promoted more than conventional agriculture.
18. Food should only be allowed to be imported if it meets Swiss standards of production regarding sustainability, animal rights, and workers' rights.

19. There should be a universal basic income in Switzerland.

20. The state should offer more resources for the construction of social housing.

21. There should be an upper limit to foreign students at Swiss universities.

22. University studies should be strictly free of charge.

C Political statement descriptives

Table 1: Cohort 1, means, standard deviations, and valid responses to policy items

	Wave 1			Wave 5		
Item nr.	N	M	SD	N	M	SD
1	154	3.35	1.95	118	3.64	1.85
2	154	3.62	1.76	118	3.67	1.79
3	154	5.60	1.51	118	5.42	1.57
4	154	5.41	1.67	118	5.51	1.63
5	154	3.29	1.52	118	3.19	1.49
6	154	2.91	1.62	118	2.94	1.66
7	153	5.44	1.43	118	5.45	1.36
8	154	4.32	1.76	118	4.34	1.73
9	153	4.98	1.82	118	4.74	1.81
10	154	4.23	1.57	118	4.11	1.55
11	154	3.13	1.53	118	3.27	1.53
12	153	3.44	1.63	118	3.41	1.64
13	154	3.68	1.68	118	2.69	1.58
14	154	5.86	1.27	118	5.76	1.30
15	154	5.51	1.30	118	5.40	1.33
16	153	5.39	1.38	118	5.28	1.42
17	154	5.36	1.26	118	5.32	1.31
18	154	5.34	1.23	118	5.45	1.17
19	154	3.88	1.89	118	3.81	1.82
20	154	4.31	1.42	118	4.62	1.47
21	154	2.44	1.46	118	2.58	1.71
22	154	4.69	1.75	118	4.76	1.77

Note. Ranges from 1 (strongly disagree) to 7 (strongly agree)

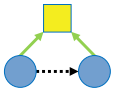
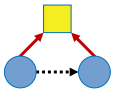
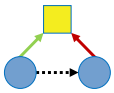
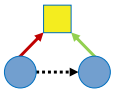
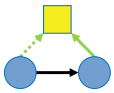
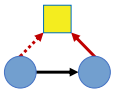
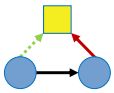
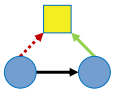
Table 2: Cohort 2, means, standard deviations, and valid responses to policy items

	Wave 1			Wave 5		
Item nr.	N	M	SD	N	M	SD
1	393	3.77	1.83	272	4.00	1.73
2	391	3.83	1.59	272	3.93	1.61
3	393	5.50	1.63	272	5.18	1.75
4	392	5.56	1.56	272	5.38	1.61
5	392	2.84	1.53	271	2.79	1.43
6	393	2.89	1.58	270	2.77	1.49
7	393	4.82	1.64	272	4.89	1.50
8	390	3.79	1.84	272	3.80	1.69
9	393	4.77	1.86	272	4.77	1.76
10	393	3.94	1.56	272	3.77	1.48
11	391	3.25	1.67	272	3.38	1.66
12	392	3.65	1.71	271	3.54	1.70
13	392	3.62	1.68	272	3.24	1.69
14	391	5.84	1.36	272	5.62	1.41
15	393	5.47	1.52	272	5.30	1.48
16	393	5.21	1.54	272	5.18	1.51
17	392	5.20	1.32	272	5.12	1.35
18	393	5.11	1.45	272	5.16	1.43
19	393	3.40	1.86	273	3.32	1.74
20	391	4.19	1.41	273	4.26	1.38
21	392	2.59	1.61	273	3.01	1.59
22	393	4.72	1.74	273	4.56	1.80

Note. Ranges from 1 (strongly disagree) to 7 (strongly agree)

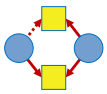
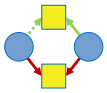
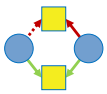
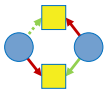
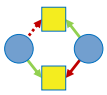
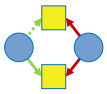
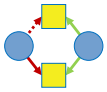
D Main SAOM results

Table 3: Main Stochastic Actor Oriented Model parameter estimates and standard errors: Selection and influence

		Cohort 1	Cohort 2
	Homophily from pos. att.	0.05* (0.02)	0.02 (0.02)
	Homophily from neg. att.	0.07* (0.03)	0.06* (0.03)
	Heterophily type 1[a]	−0.05 (0.05)	0.05 (0.04)
	Heterophily type 2[a]	−0.04 (0.05)	0.03 (0.03)
	Influence, pos. att.	0.06* (0.02)	−0.01 (0.02)
	Influence, neg. att.	0.06 (0.03)	−0.01 (0.03)
	Influence, pos. att. and neg. att.	−0.11* (0.05)	−0.02 (0.04)
	Influence, neg. att. and pos. att.	0.08 (0.05)	−0.07 (0.04)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors in parentheses. Dependent tie is dotted line.
[a] Newly programmed effect.

Table 4: Main Stochastic Actor Oriented Model parameter estimates and standard errors: Convergence and divergence

		Cohort 1	Cohort 2
	Latent convergence type 1, pos. att.[b]	0.31*** (0.03)	0.13*** (0.01)
	Latent convergence type 1, neg. att.[b]	0.63*** (0.07)	0.26*** (0.02)
	Latent convergence type 2, pos. att.[b]	0.07 (0.05)	0.01 (0.02)
	Latent convergence type 2, neg. att.[b]	0.18*** (0.05)	0.07*** (0.02)
	Latent divergence type 1, pos. att.[ab]	0.87*** (0.12)	0.38*** (0.04)
	Latent divergence type 1, neg. att.[ab]	0.66*** (0.09)	0.31*** (0.03)
	Latent divergence type 2, pos. att.[ab]	0.57*** (0.12)	0.37*** (0.04)
	Latent divergence type 2, neg. att.[ab]	0.47** (0.15)	0.36*** (0.05)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors in parentheses. Dependent tie is dotted line.
[a]Newly programmed effect. [b]Estimate and SE multiplied by 100.

E Interpretation of auxiliary results of the main model

For the results described below, see the tables at the end of this section. The rates in the friendship networks ranged from 4.30 to 6.97 for both cohorts. In the political attitude network, rates ranged from 3.55 to 7.37. This means that on average, individuals are estimated to have considered changing one of their outgoing friendship ties and attitude ties in each network around 4 to 7 times between subsequent observations.

Examining endogeneity in the friendship networks of both cohorts, we find positive and significant effects of reciprocity and transitivity. Furthermore, their interaction was negative and significant, as shown previously by Block (2015). Indegree popularity was positive and significant, suggesting individuals prefer to connect to alters who have more incoming friendship ties. The related effects of outdegree popularity and outdegree activity were both negative and significant, suggesting that individuals prefer to connect to alters who send fewer friendship nominations, and that they prefer to nominate fewer people as friends as well.

Turning to homophily, we find positive and significant effects of speaking the same language, being from the same study major (in Cohort 1), and gender. A positive, yet non-significant effect for age homophily was found in Cohort 2. In Cohort 2 there were additional positive significant effects of gender on friendship, with women more likely to nominate, and be nominated as friends than men. Indegree popularity for positive attitude about a political statement was positive significant in both samples, but not significant in either sample for negative attitude. The effect of many positive attitudes on others' negative attitude on a statement and vice versa was negative and significant, but only in Cohort 2.

Table 5: Full Stochastic Actor Oriented Model parameter estimates.

	Cohort 1	Cohort 2
Rate friend W2-W3	6.87 (0.55)***	5.98 (0.36)***
Rate friend W3-W4	6.97 (0.58)***	6.48 (0.36)***
Rate friend W4-W5	4.30 (0.37)***	4.94 (0.28)***
Density friend (intercept)	−3.10 (0.15)***	−3.57 (0.10)***
friend: reciprocity	3.49 (0.18)***	3.85 (0.14)***
friend: transitive triplets	0.81 (0.05)***	1.22 (0.04)***
friend: transitive recipr. triplets	−0.62 (0.09)***	−1.12 (0.07)***
friend: indegree - popularity	0.07 (0.03)**	0.06 (0.01)***
friend: outdegree - popularity	−0.18 (0.03)***	−0.20 (0.02)***
friend: outdegree - activity	−0.02 (0.01)*	−0.01 (0.01)*
friend: anti-isolate	−1.70 (0.39)***	−1.41 (0.35)***
friend: ethnicity	0.23 (0.08)**	0.33 (0.06)***
friend: same major	0.50 (0.06)***	
friend: gender alter	0.05 (0.07)	0.27 (0.07)***
friend: gender ego	−0.16 (0.08)	0.49 (0.08)***
friend: same gender	0.15 (0.07)*	0.42 (0.07)***
friend: birthyear similarity	0.44 (0.29)	0.45 (0.22)*
friend: neg.att-pos.att two-paths^{a}	−0.05 (0.05)	0.05 (0.04)
friend: pos.att-neg.att two-paths^{a}	−0.04 (0.05)	0.03 (0.03)
friend: pos.att agreement	0.05 (0.02)*	0.02 (0.02)
friend: neg.att agreement	0.07 (0.03)*	0.06 (0.03)*
Overall max. t-ratio	0.19	0.14
Param. max. t-ratio	0.04	0.04
Iterations	8147	8147

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Dependent tie is dotted line. ^{a}Newly programmed effect. ^{b}Estimate and SE multiplied by 100

Table 6: Full Stochastic Actor Oriented Model parameter estimates. (cont)

	Cohort 1	Cohort 2
Rate pos W2-W3	6.30 (0.38)***	7.32 (0.33)***
Rate neg W2-W3	4.79 (0.35)***	5.43 (0.27)***
Rate pos W3-W4	7.37 (0.46)***	7.12 (0.30)***
Rate neg W3-W4	4.97 (0.39)***	5.46 (0.27)***
Rate pos W4-W5	6.52 (0.43)***	6.73 (0.29)***
Rate neg W4-W5	3.55 (0.28)***	4.31 (0.22)***
pos.att: outdegree (density)	−1.53 (0.12)***	−1.47 (0.07)***
neg.att: outdegree (density)	−1.75 (0.15)***	−1.61 (0.08)***
pos.att: friend to agreement	0.06 (0.02)*	−0.01 (0.02)
neg.att: friend to agreement	0.06 (0.03)	−0.01 (0.03)
pos.att: friend and neg.att two-paths^{a}	−0.11 (0.05)*	−0.02 (0.04)
neg.att: friend and pos.att two-paths^{a}	0.08 (0.05)	−0.07 (0.04)
pos.att: 4-cycles (1)^{b}	0.31 (0.03)***	0.13 (0.01)***
neg.att: 4-cycles (1)^{b}	0.63 (0.07)***	0.26 (0.02)***
pos.att: shared neg.att (1) to agreement^{b}	0.07 (0.05)	0.01 (0.02)
neg.att: shared pos.att (1) to agreement^{b}	0.18 (0.05)***	0.07 (0.02)***
pos.att: indegree - popularity^{b}	0.65 (0.20)**	0.19 (0.06)**
neg.att: indegree - popularity^{b}	0.14 (0.37)	0.06 (0.09)
pos.att: indegree neg.att popularity^{b}	−0.27 (0.31)	−0.57 (0.09)***
neg.att: indegree pos.att popularity^{b}	−1.02 (0.40)**	−0.46 (0.09)***
pos.att: different neg.att (1) to disagreement type 2^{ab}	0.57 (0.12)***	0.37 (0.04)***
neg.att: different pos.att (1) to disagreement type 2^{ab}	0.47 (0.15)**	0.36 (0.05)***
Overall max. t-ratio	0.19	0.14
Param. max. t-ratio	0.04	0.04
Iterations	8147	8147

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Dependent tie is dotted line. ^{a}Newly programmed effect. ^{b}Estimate and SE multiplied by 100

F Summary of effect and goodness of fit statistics newly added to RSiena

F.1 Effects

SAOM effects for three networks with mixed two-mode and one-mode nodesets were not implemented in the RSiena software at time of writing. Thus, four new effects were added to RSiena for the purposes of our study. First of these were two, three-network, mixed triangle effects. These are (using the conventional RSiena shortnames) the "fromDiff" and "toDiff" effects, respectively, in analogy to "from" and "to" effects for two first mode nodes having a common bipartite node to which they connect in the same network type, already present in RSiena.

For the "fromDiff" effect, the dependent tie is the first-mode tie between two actors, dependent on their having a different connection type to a second-mode node they have in common. These correspond to the parameters reported in Table 3 as "Heterophily type 1" and "Heterophily type 2".

For the "toDiff" effect, the dependent tie is a second-mode tie of a different type than the one held by another actor to which the first actor has an outgoing tie in the first mode. This effect was used in the estimation of parameters indicated by "Influence, pos. att and neg. att" and "Influence, neg. att and pos. att".

In language more familiar to Siena users: "fromDiff" is the effect of the mixed 2-instar to second-mode node m from i in network W and from j in network V, on the directed tie from i to j in network X.

The "toDiff" effect is the effect of the directed twopath from actor i to node m, via first-mode node j with ties in network X and W respectively, on actor i's connection to m in network V.

The second two new effects relate to the mixed four cycles in the bipartite networks - in particular the oppositional kinds. These are named "oppCycle4.1" and "oppCycle4.2" effects. The .1 and .2 indicate that these are two different forms of oppositional four-cycles, as for our purposes we saw no reason to distinguish between the two conceptually. For both of the statistics below, X represents the dependent network, W the independent.

oppCycle4.1 is the clockwise path of ties $XWWX$, starting at an actor sending tie in network X to the second mode node. This effect was used to estimate parameters reported in Table 1 as "Latent divergence type 2, pos. att." and "Latent divergence type 2, neg. att.".

oppCycle4.2 is the clockwise path of ties $XWXW$, with the same starting position. This effect was used to estimate parameters reported in Table 1 as "Latent divergence type 1, pos. att." and "Latent divergence type 1, neg. att.".

F.2 Auxiliary statistics (Goodness of fit)

We added two kinds of new auxiliary statistics, which we used for goodness-of-fit tests in SAOMs. First, we defined a census of mixed triads of three nodes, two first-mode and two second-mode, including up to three networks. Ties can

only be reciprocal in the pair of first-mode nodes. Additionally, the bipartite networks are disjoint (i.e. there cannot be a tie present in both networks, for one dyad at the same time). This census is in the exclusive form, i.e. a closed triad is not additionally counted in any separate twopath configuration. Similarly, we count a reciprocal tie between two nodes only once.

Secondly, we created a new census of complete (i.e. where all pairs of nodes which has one in the second mode and one in the first mode set have a connection) four-cycles for configurations of two first-mode nodes, two second-mode nodes, and two network types connecting them, which are again disjoint.

G Network descriptives

Friendship descriptives

Table 7: Friendship network descriptives, Cohort 1

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
N	202	236	225	215	210
Structural missing	59	25	36	46	51
Density	0.01	0.02	0.02	0.02	0.02
Mean degree	0.92	2.39	2.30	2.44	1.92
SD indegree	1.26	2.40	2.37	2.59	2.10
SD outdegree	1.73	3.17	3.08	3.29	2.69
Jaccard (t vs. $t-1$)		0.37	0.66	0.68	0.75

Structural missing indicates individuals not part of the cohort at time of data collection.

Table 8: Friendship network descriptives, Cohort 2

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
N	529	635	625	620	603
Structural missing	131	25	35	40	57
Density	0.005	0.01	0.01	0.01	0.01
Mean degree	1.54	1.93	2.22	2.55	2.15
SD indegree	1.65	1.90	2.16	2.22	2.13
SD outdegree	1.98	2.63	2.94	3.17	3.31
Jaccard (t vs. $t-1$)		0.55	0.68	0.70	0.76

Structural missing indicates individuals not part of the cohort at time of data collection.

Attitudinal descriptives

Table 9: Political network descriptives, Cohort 1

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Density, pos	0.37	0.34	0.34	0.34	0.34
Density, neg	0.23	0.24	0.21	0.23	0.22
Mean indegree, pos	49.55	44.27	42.73	42.23	36.18
Mean indegree, neg	30.59	31.41	26.36	28.91	23.91
Mean outdegree, pos	4.18	3.73	3.60	3.56	3.05
Mean outdegree, neg	2.58	2.65	2.22	2.44	2.02
SD indegree, pos	33.69	24.94	25.82	26.39	23.20
SD indegree, neg	25.69	24.06	22.40	22.79	20.41
SD outdegree, pos	4.41	4.25	4.39	4.31	4.20
SD outdegree, neg	3.00	3.22	2.99	3.19	2.94
Jaccard, pos (t vs. $t-1$)		0.76	0.76	0.76	0.77
Jaccard, neg (t vs. $t-1$)		0.73	0.72	0.72	0.78

Table 10: Political network descriptives, Cohort 2

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Density, pos	0.35	0.32	0.31	0.30	0.31
Density, neg	0.25	0.24	0.23	0.24	0.23
Mean indegree, pos	119.27	88	91.32	90.64	76.23
Mean indegree, neg	84.64	66.27	67.68	72.09	56.95
Mean outdegree, pos	3.98	2.93	3.04	3.02	2.54
Mean outdegree, neg	2.82	2.21	2.26	2.40	1.90
SD indegree, pos	81.51	55.40	58.57	56.53	52.26
SD indegree, neg	62.74	45.90	47.04	50.54	41.11
SD outdegree, pos	4.12	3.89	3.88	3.77	3.66
SD outdegree, neg	3.28	3.30	3.25	3.33	3.10
Jaccard, pos (t vs. $t-1$)		0.72	0.71	0.73	0.72
Jaccard, neg (t vs. $t-1$)		0.72	0.72	0.69	0.73

H Main GOF statistics aggregated across waves

Figures 1 and 2 show the most relevant structures which the estimated model should reproduce. These are those that are also used in the calculation of network polarization. In these figures, the x-axis labels give a notational and graphical representation of all relevant tetradic and triadic multilevel network structures.

The violin plots here show the frequency distribution of the structure specified on the x-axes, in networks generated by simulating from each observation wave until the following observation 5000 times, using the estimated model. The red diamond represents the observed value relative to these simulations, aggregated across all simulations. Numbers are the true counts of the given statistic. A better fitting model has observed values closer to the center of the simulated distribution. Using the Mahalanobis distance of the given structures in the observed versus simulated data, we test for significant discrepancy between our model and our observation. If the p-value is low, observations in the data fall too far to the extremes of the distributions generated by model simulation and therefore it is unlikely that our observed data could have been drawn from the estimated model.

First, we examine the mixed four cycles of pairs of individuals and their attitudes, indicating the attitude clustering regardless of social connection, as given in Formulas 1 and 2 in the main text. Figure 1 shows that the model plausibly reproduces the observed data in terms of attitudes between individuals in both cohorts, although there is a tendency towards over-generation of perfect disagreement structures and mixed agreement-disagreement structures in Cohort 2. Overall, the model fits well to the focal features. Goodness of fit on other features of the network are reported in Appendix J.

Second, we examine mixed triads of individuals and political statements, connected by friendships and attitudes. As mentioned previously, these represent the clustering of individuals with shared (or opposing) attitudes, i.e. the social aspect of polarization given in Formulas 3 and 4. Figure 2 shows the fit of the model to the observed census of mixed triads in the network. We see that the model does not significantly misfit on these structures in either cohort. However, it seems that in Cohort 1 there are slightly too many structures with reciprocal friendship ties generated by the model, and slightly too few with one-sided friendships. This bias is not apparent in Cohort 2.

In analyses not reported here, separation of the fit statistics by period (i.e. counting the structures at each sequential wave, instead of in aggregate) suggests the model significantly overestimates the tendency towards specific polarization structures at the earlier time point, but underestimates them at the later one, particularly in Cohort 2 (Appendix I). Given the high number of parameters included in the model, however, we opt to retain the uniform parameter estimates to reduce the risk of overfitting and type II errors.

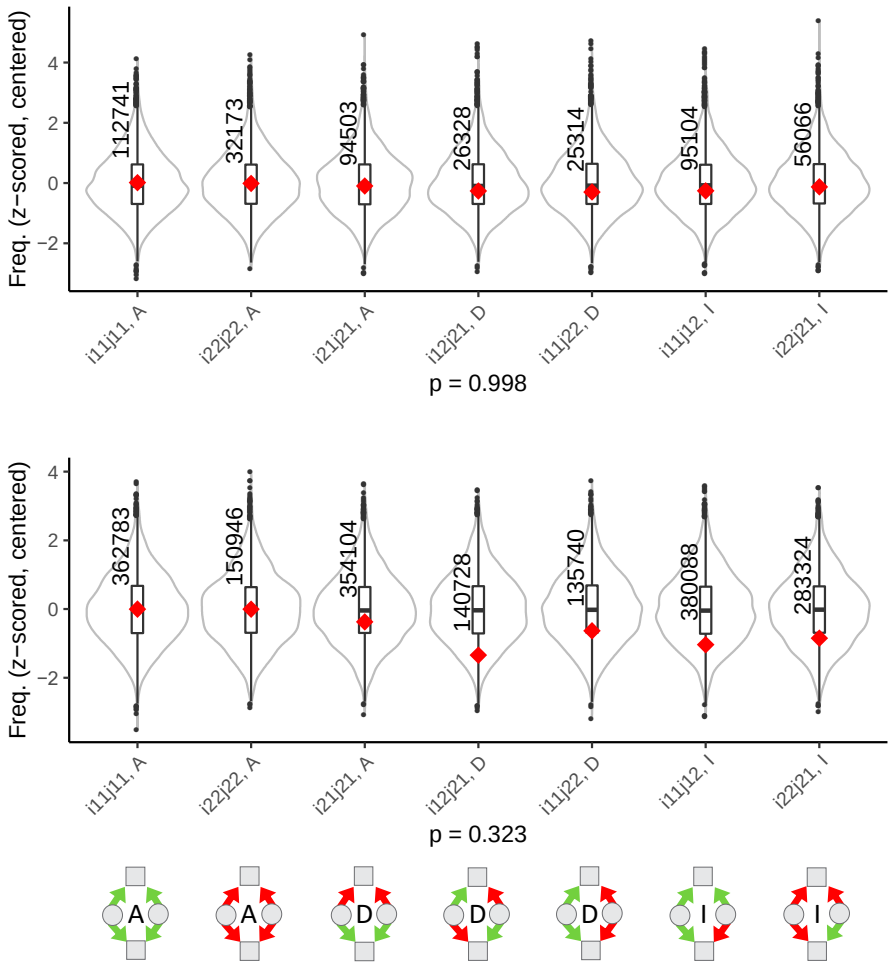


Figure 1: Goodness-of-fit on four-cycle meso-structures. Upper row is Cohort 1, lower row is Cohort 2. Columns represent structures with corresponding notation and pictogram. i and j represent the individuals on the left and right, with the first digit representing the upper tie and the second digit the lower. Digits 1 and 2 indicate positive and negative network. p-values from Mahalanobis distance between all structures in simulation and observed data.

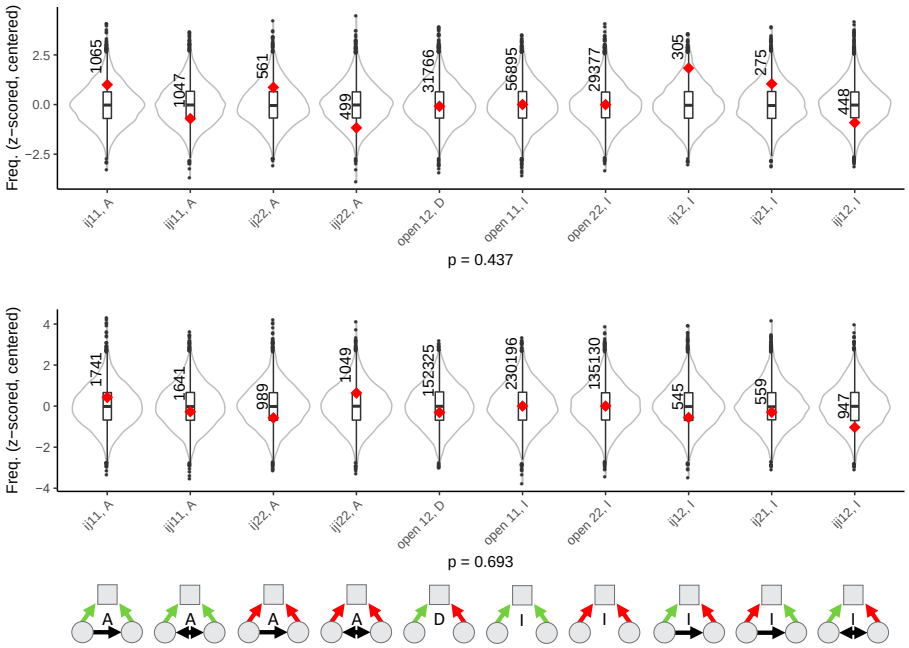


Figure 2: Goodness-of-fit on triadic meso-structures. Upper row is Cohort 1, lower row is Cohort 2. Columns represent structures with corresponding notation and pictogram. *ij* and *iji* represent single and reciprocal social ties respectively. Digits 1 and 2 indicate positive and negative attitude network. First position indicates the tie of the individual on the left, the second, the individual on the right. *p*-values from Mahalanobis distance between all structures in simulation and observed data.

I Main GOF statistics, by wave

Goodness-of-fit statistics for the main models that immediately relate to the outcome of network polarization are shown in Figure 3. These were developed as part of the project and are now available for use of other researchers. They assess the number of mixed triads and four-cycles that are part of the newly proposed network polarization metric and its underlying census. An overview of the network structures was shown in Fig.1 in the main text. Numbers 1 and 2 in the x-axis refer to positive and negative ties, respectively. The number of mixed triads is 10 instead of 7 as for each of the closed trias, also the structure with reciprocal ties (marked as iji in the x axis) is considered. The final two structures are equivalent when the social tie is reciprocated. In an aggregated form, these GOF statistics are the basis for the micro-macro assessment in section 6.3 of the main article.

J Other goodness of fit statistics for main models

Figures 4–6 show additional goodness-of-fit statistics for the main models. These are standard GOF statistics and are available in the RSiena Software (Ripley et al., 2019). They measure to what degree the simulated networks are similar with regards to truad census and degree distributions.

K Robustness checks

K.1 Model results

While we believe the best model and data specification is the main model presented, we conducted five robustness checks. Results from all of these models can be found in the supplemental HTML file all_saoms_full_results.html. Results summaries follow below.

In text, we refer to each variant by codes R1-R5. Four of the alternative specifications were based on data modifications:

First, since our data extended beyond the one-year period used in the main analysis, we additionally tested until the end of the students' second year of their undergraduate studies (R1; i.e. nine waves of data). This was not included in the main analysis because of substantial change in the size and particularly participation of the cohorts after the end of the first year. Participation dropped from 60% at Wave 5 to 48% at Wave 6 for Cohort 1, and from 48% to 36% in the same period for Cohort 2 [reference removed for review].

In a second test, we estimated the model for the first to second observation (R2). The first wave of data had been initially excluded due to difficulties it caused in estimating the model, likely due to the high instability of the friendship network from Wave 1 to 2 in both cohorts.

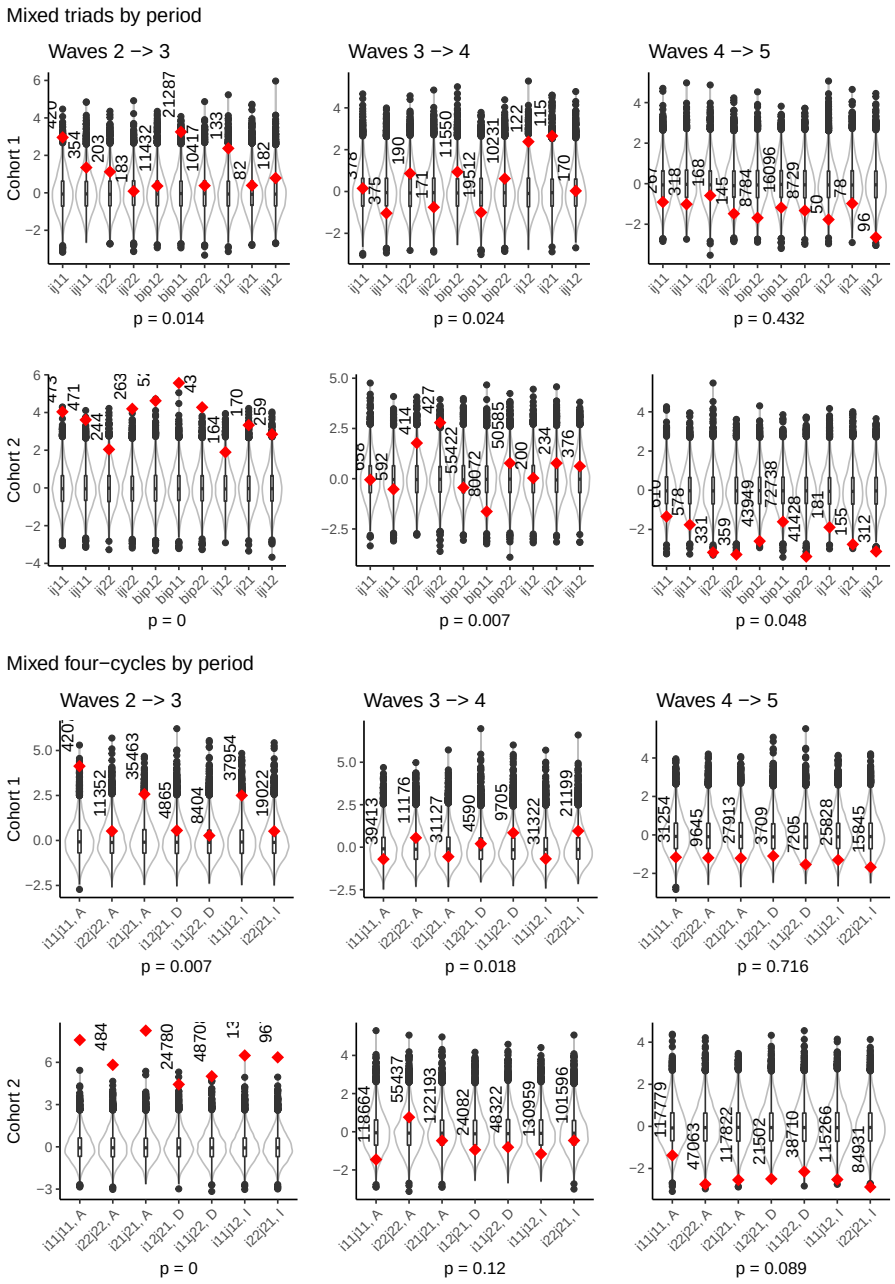


Figure 3: Main GOF statistics, by wave

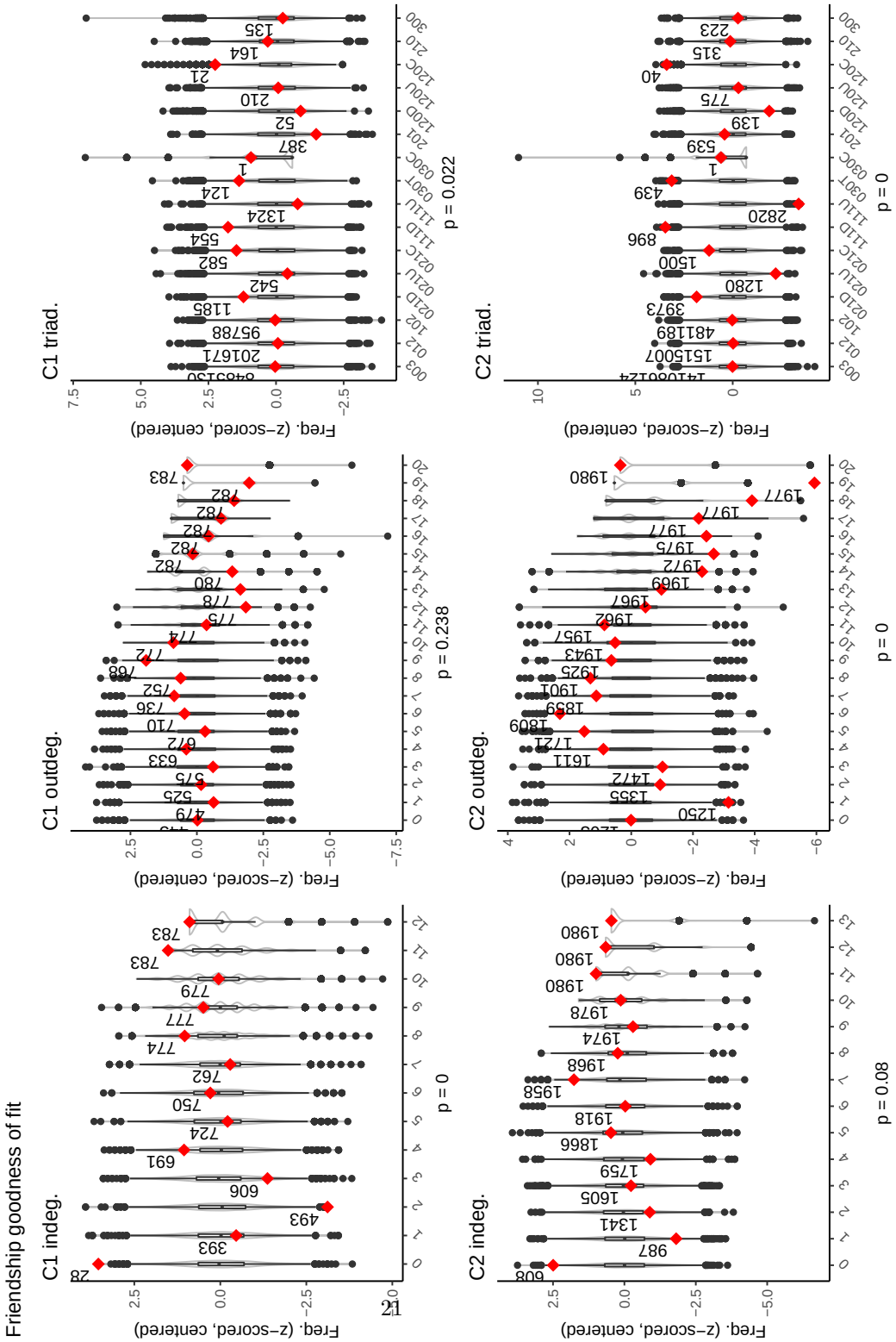


Figure 4: Other goodness of fit statistics for main models part 1

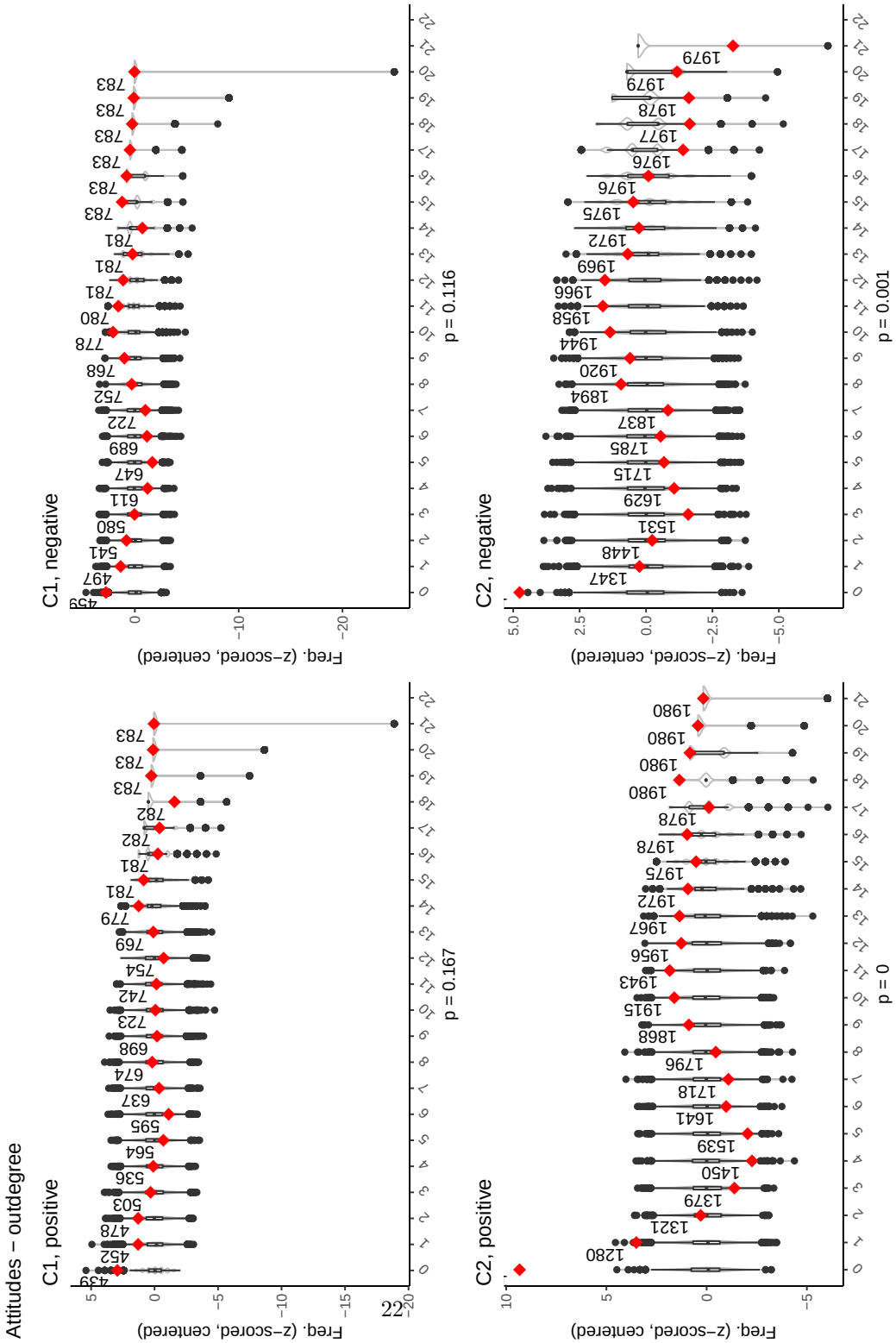


Figure 5: Other goodness of fit statistics for main models part 2

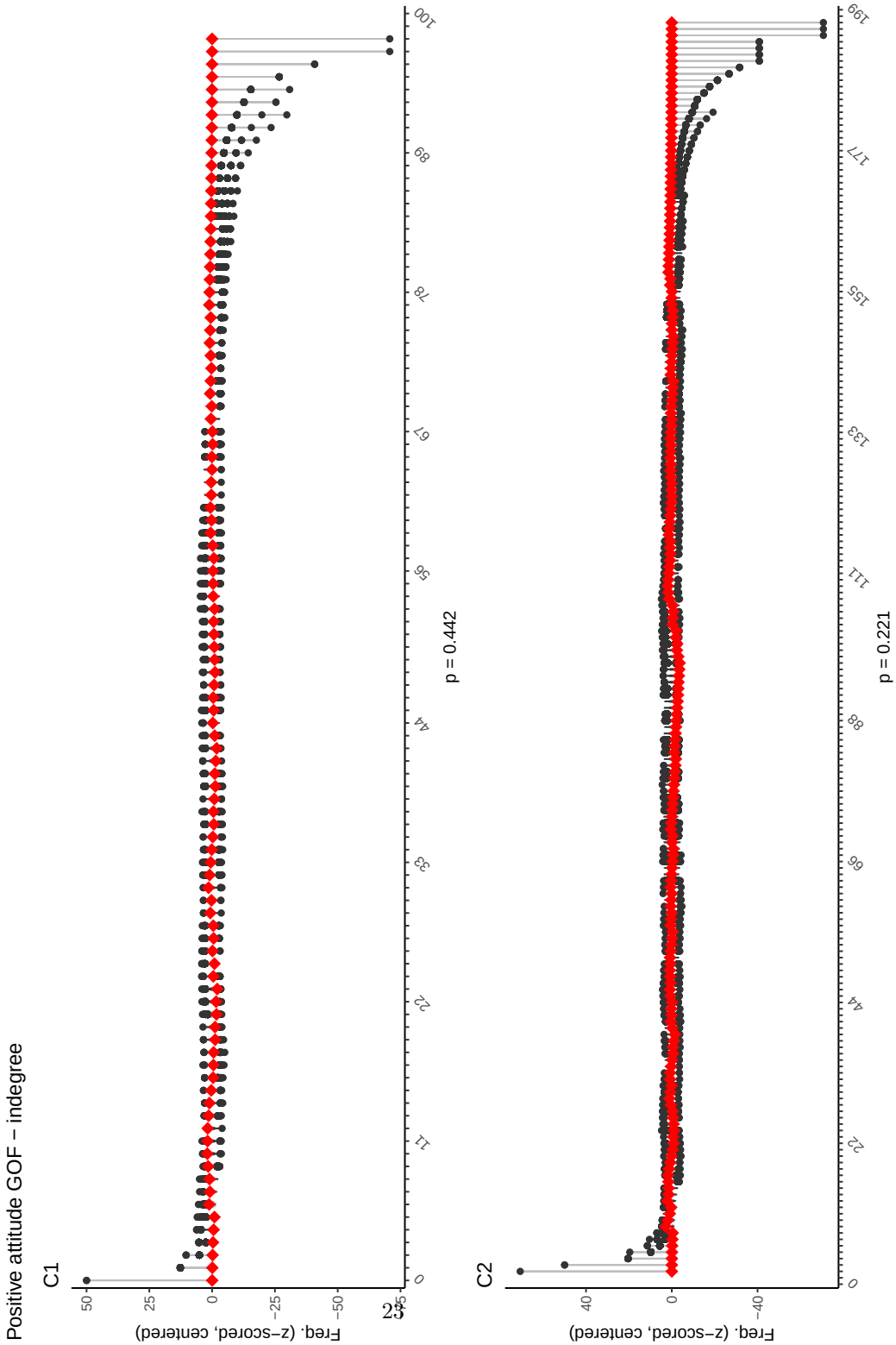


Figure 6: Other goodness of fit statistics for main models part 3

Third, we estimated the model with a response to an attitude item of "no opinion" as missing, rather than a neutral attitude (R3). Notably, this model misfit significantly on the four-cycles examining the ideological part of polarization ($p = .037$) for cohort two (tending to generate more four-cycles than observed).

Fourth, we estimated models with different thresholds for coding ties in the attitude network (R4). We considered the alternatives one point away from the neutral option on the scale (i.e. a less extreme threshold) and those at the outer ends of the scales (i.e. a more extreme threshold). The former of these two models did not converge, likely due to minimal tie changes – most people who held a valenced attitude on a topic either maintained an attitude of that valence, becoming less extreme or becoming neutral, rather than switching to the other side.

Lastly, focusing on alternative model specification, and in line with previous literature (Wang et al., 2020; Lazer et al., 2010), we considered the effect of political orientation as a univariate dimension predicting social ties which may correlate with other political attitudes (R5). We thus included a homophily term for a simple left-right political orientation scale.

As with the main results, we interpret both the model parameters themselves, as well as multivariate Wald tests presented in the table at the end of this section, which test the sum of all parameters representing a single hypothesis. This results table is followed by goodness-of-fit plots on the focal structures.

Regarding H1a, that people will tend to have friendships with others with whom they share political attitudes, results are largely consistent with the main model. One of the two parameters was consistently positive and significant in Cohort 1, and never significant in Cohort 2, as in the main model. The second of the two, which was positive and significant in the main model for both cohorts, was not significant in either cohort in R2 or R4 (i.e. in the first period or with a higher binarization threshold). In R1, R5 and R3, the parameter was reduced to non-significance for Cohort 1. For Cohort 2, the parameter remained positive and significant for R1, R3 and R5. Using Wald tests grouping the parameters, results are significant except for Cohort 2 R2 and R4, though the χ^2 statistic tends to decrease relative to the main model.

H1b, that people will tend to avoid friendships with people with opposed attitudes, was consistent with the main model; not supported with either parameter in any robustness check model. Similarly, Wald tests of the grouped parameters testing this hypothesis found no significant effect of selection. The majority of the models contained at least one opposite-expectation parameter sign.

H2a, that people will tend to hold attitudes that their friends hold, was largely consistent with the main model, with some exceptions. As in the main model, a positive and significant effect was found for influence on positive attitudes towards a topic in Cohort 1 in R1 and R5, but no significant effect in R2, R3, and R4. In contrast, for Cohort 2, the expected effect was positive and significant in R1 and R4, unlike the non-significant effect in the main model. For the second effect testing H2a on negative attitudes, the positive non-significant

effect in the main model for Cohort 1 was significant in R1, but not in any other model. For Cohort 2, which did not have a significant parameter in the main model, it remained unsupported in other models. Wald tests supported the hypothesis in both cohorts for R1 (as opposed to the main model where the hypothesis was only supported in Cohort 1). In R4 and R5, the hypothesis was supported with a similar pattern to the main model, i.e. only in Cohort 1. In R2 and R3 the hypothesis was not supported by the Wald tests.

H2b, that people will tend to avoid holding attitudes that are opposed to their friends', was tested by two parameters. For the first of these, the robustness check models replicated the positive significant effect in Cohort 1 only in R5. Consistent with the findings for Cohort 2, no effect was found in any of the robustness check models. For the second parameter, the non-significant parameter found in Cohort 1 was increased to positive significance (counter to expectation) in R3, but was non-significant in R1, R2, R4, and R5. The negative parameter found in Cohort 2 became significant in R1, R3 and R4, in line with expectation, while staying non-significant in R5 and R2. Wald tests yield results similar to the main model: Only in the case of R1 is the test of the parameters significant, and only for Cohort 2.

Moving to H3a, that individuals sharing some pre-existing attitude are more likely to share more attitudes in the future, results are more straightforward. Of the first two of four parameters, both remained positive and significant in all robustness checks for both cohorts, as in the main model. The third parameter, which was not significant in either cohort in the main model, became positive significant in R2 for Cohort 1 and non-significant in R3 for Cohort 2. The fourth parameter, positive and significant for both cohorts in the main model, was supported fully in R1 and R5. In R2 and R3 it was only supported for Cohort 1, and In R4 for Cohort 2. Wald tests support the hypothesis in both cohorts for all robustness checks.

For H3b, that individuals with opposing attitudes will continue to have more opposing attitudes in the future, there are again four parameters, showing robust results. In both cohorts, for all four parameters, results are the same as in the main model – parameters are positive and significant with three exceptions. Both of the the parameters become non-significant in Cohort 1 in R2, while for the same cohort in R3 one parameter becomes non-significant. Wald tests support the hypothesis in all cases.

Overall, mixed results on the hypotheses largely retain the level of support across robustness checks, although patterns are not identical. In sum, it appears that positive selection (H1a) is robustly supported, negative selection (H1b) is unsupported, positive influence (H2a) is weakly supported, negative influence (H2b) is not supported, and latent-cause similarity (H3a) and dissimilarity (H3b) are robustly supported.

Table 11: Wald tests of summed parameters

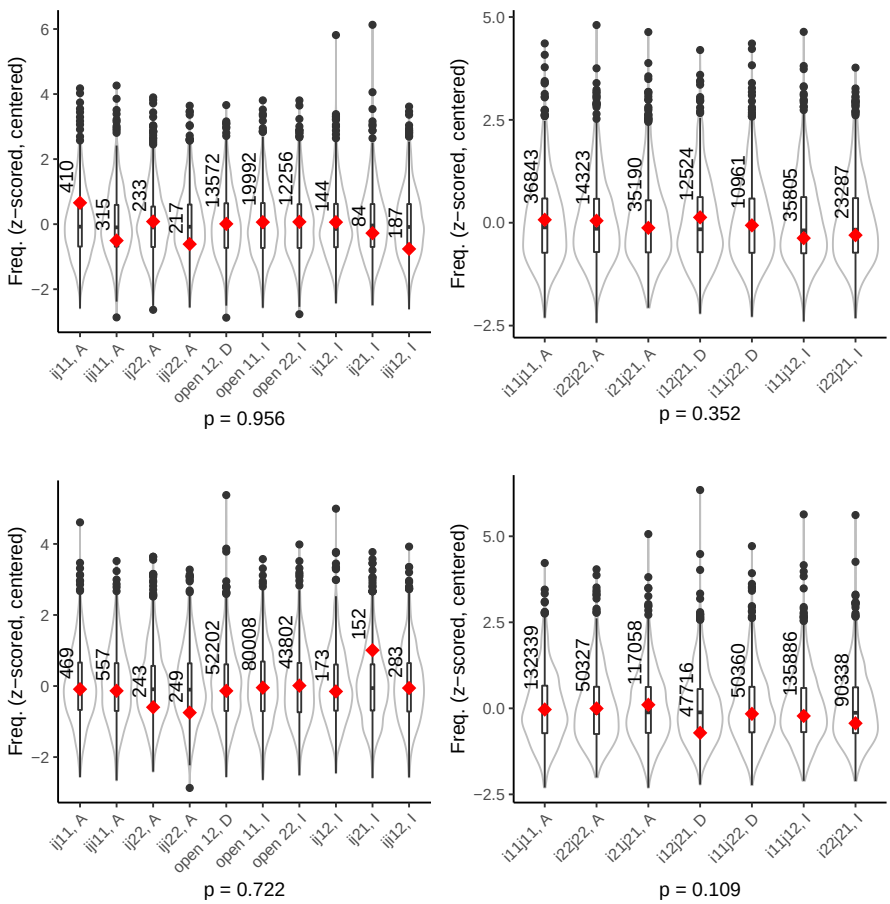
	H1a	H1b	H2a	H2b	H3a	H3b
C1, Main	13.67(1)***	1.68(1)	9.06(1)**	0.12(1)	110.65(1)***	135.28(1)***
C2, Main	5.65(1)*	*2.73(1)*	*0.39(1)*	3.29(1)	368.04(1)***	195.78(1)***
C1R1, W1-2	8.33(1)**	0.42(1)	10.43(1)**	0.02(1)	155.3(1)***	192.2(1)***
C2R1, W1-2	8.88(1)**	0.55(1)	13.16(1)***	6.63(1)*	472.93(1)***	243.48(1)***
C1R2, W2-9	4.47(1)*	0.02(1)	0.05(1)	0.37(1)	24.51(1)***	73.01(1)***
C2R2, W2-9	1.34(1)	0.03(1)	0.99(1)	2.1(1)	91.95(1)***	69.56(1)***
C1R3, no opin. as missing	10.29(1)**	0.38(1)	1.86(1)	0.68(1)	53.66(1)***	67.76(1)***
C2R3, no opin. as missing	10.63(1)**	*1.19(1)*	0.59(1)	1.65(1)	287.65(1)***	82.4(1)***
C1R4, high tie thresh.	5.47(1)*	0.26(1)	5.75(1)*	1.99(1)	64.16(1)***	4.75(1)*
C2R4, high tie thresh.	0.14(1)	*1.03(1)*	3.71(1)	2.77(1)	161.93(1)***	56.02(1)***
C1R5, pol. orientation	11.5(1)***	0.88(1)	8.6(1)**	0.1(1)	102.5(1)***	131.48(1)***
C2R5, pol. orientation	5.96(1)*	*2.11(1)*	*0.38(1)*	2.8(1)	299.88(1)***	173.52(1)***

Note. Values are χ^2 for sum of parameters with degrees of freedom in brackets. Italicized values indicate all parameters being in a direction counter to expectation. $^{*}: p<.05, ^{**}: p<.01, ^{***}: p<.001$.

K.2 Main goodness of fit statistics for robustness checks

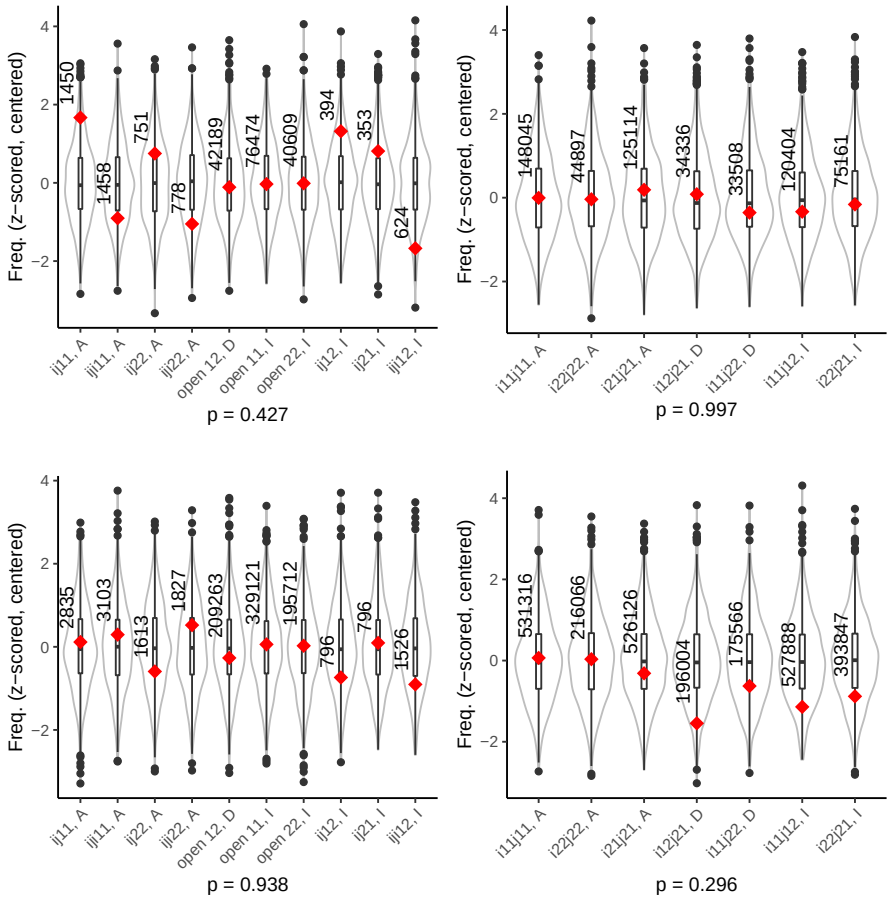
R1 (waves 1−2)

Left = triads, right = four cycles. C1 above, C2 below



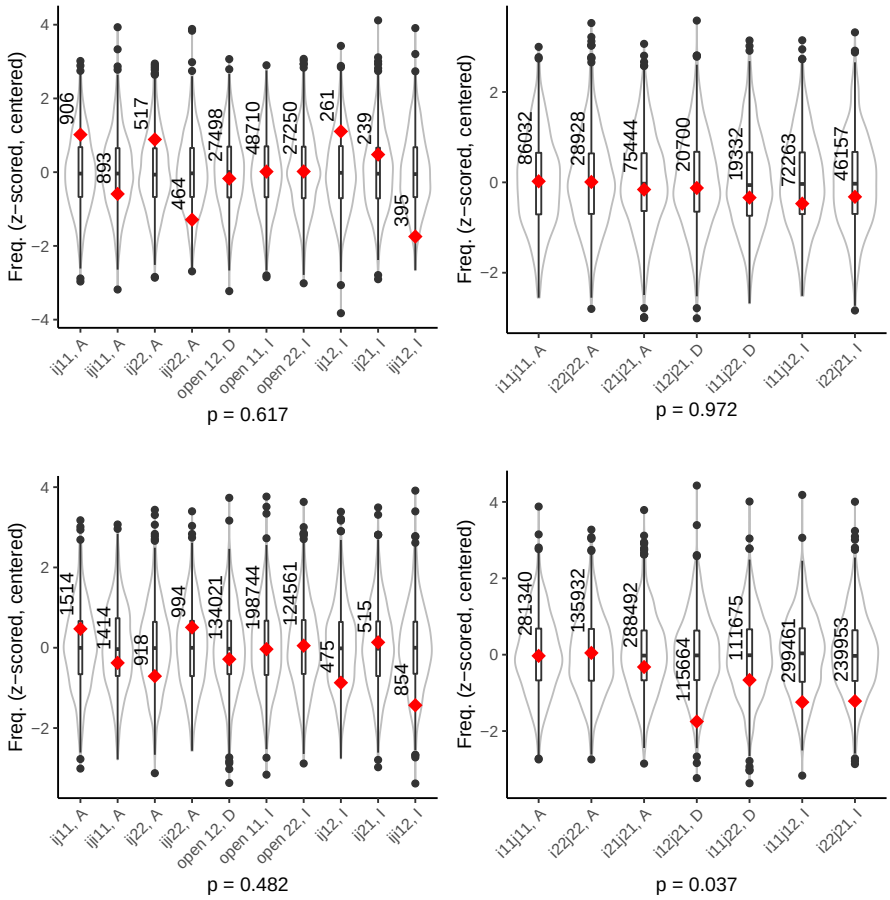
R2 (waves 2−9)

Left = triads, right = four cycles. C1 above, C2 below



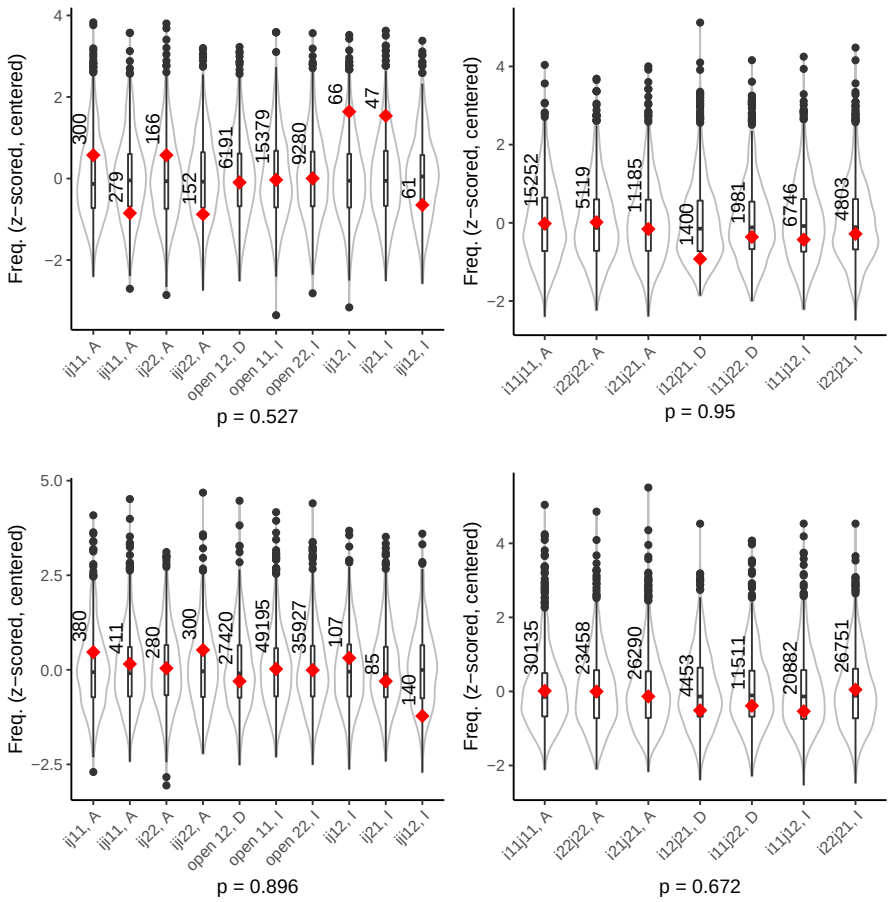
R3 (no opin as missing)

Left = triads, right = four cycles. C1 above, C2 below



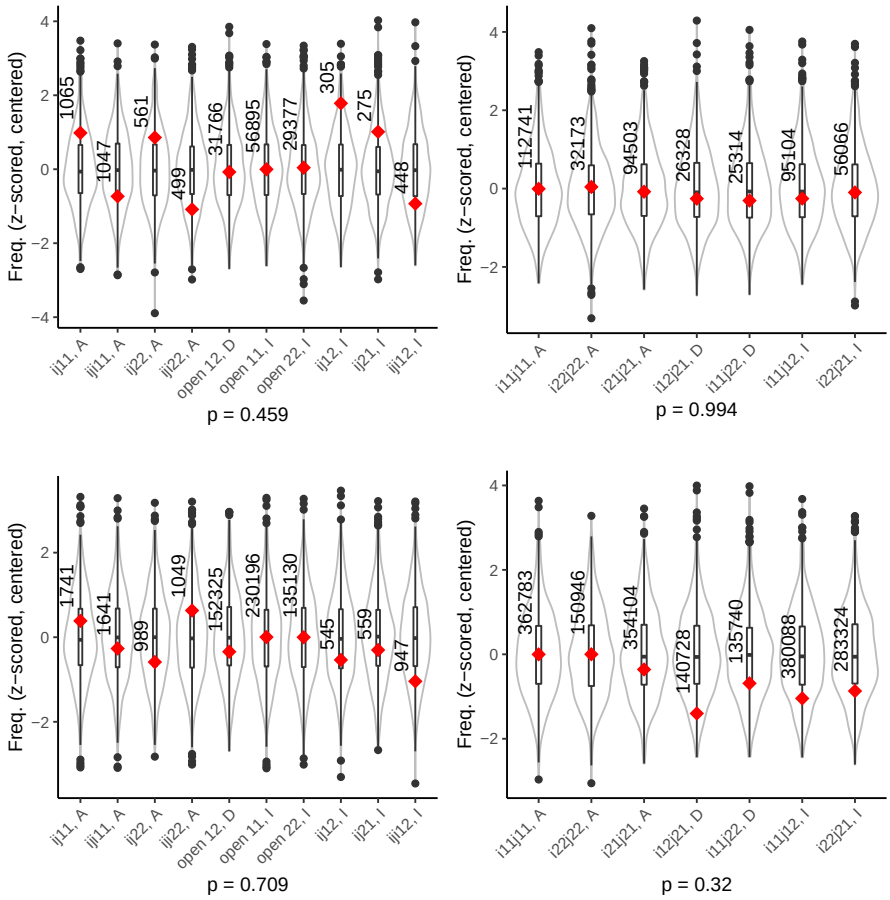
R4 (higher tie threshold)

Left = triads, right = four cycles. C1 above, C2 below



R5 (political orientation covar)

Left = triads, right = four cycles. C1 above, C2 below



L Expectation and observation by wave in four dimensions

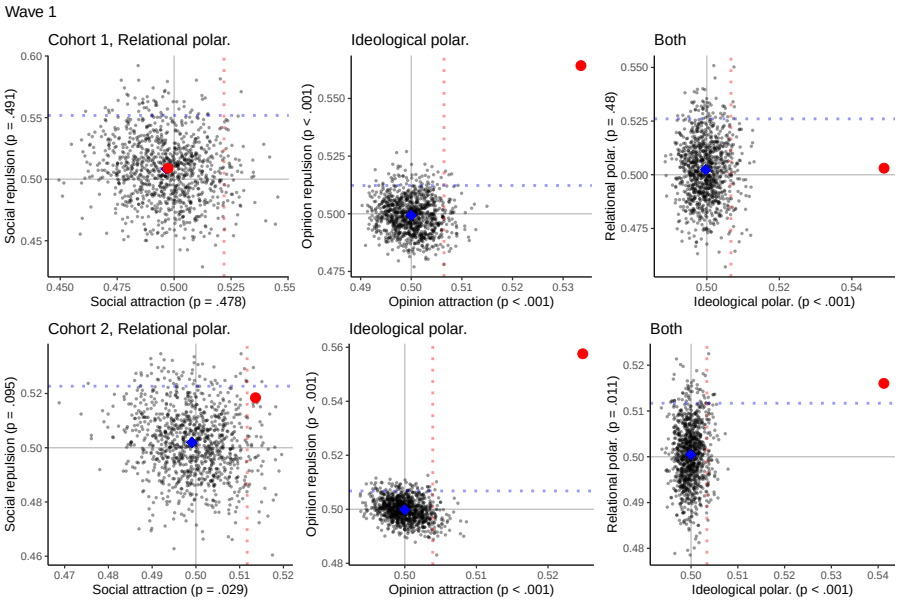
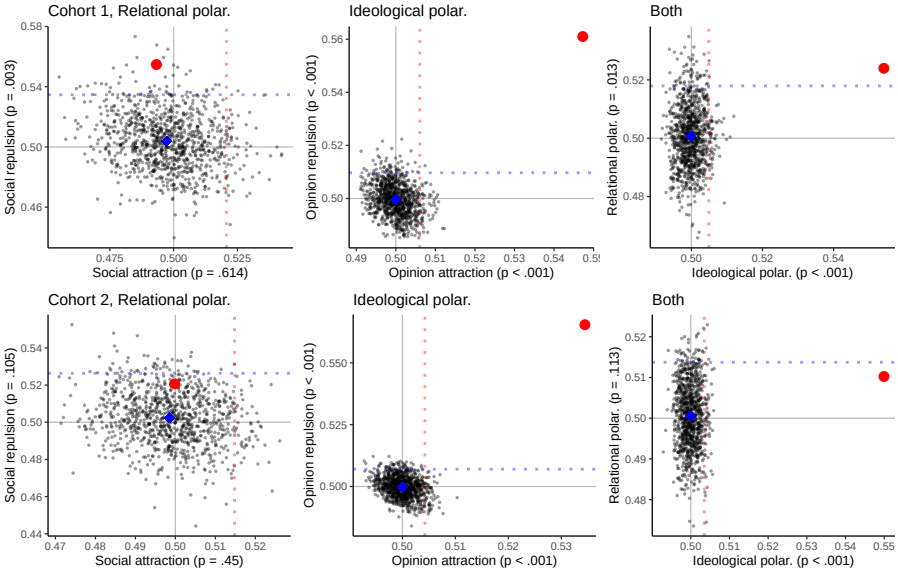
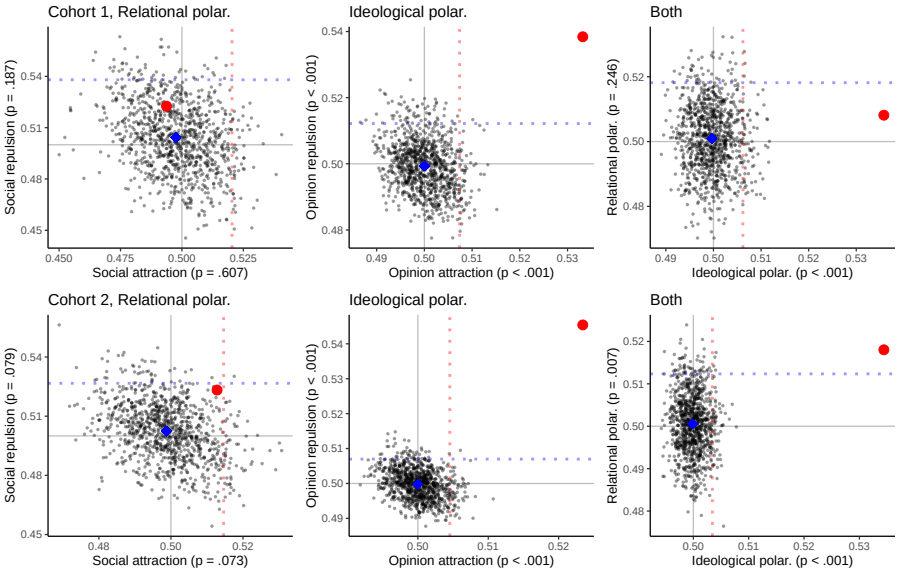


Figure 7: Two-dimensional observed network polarization (red circle) relative to expectation in two cohorts (blue diamond), with simulated values (grey dots). Dotted blue line indicates boundary for vertical axis, red indicates boundary of variable on horizontal axis, both at $p < .05$ (one-sided).

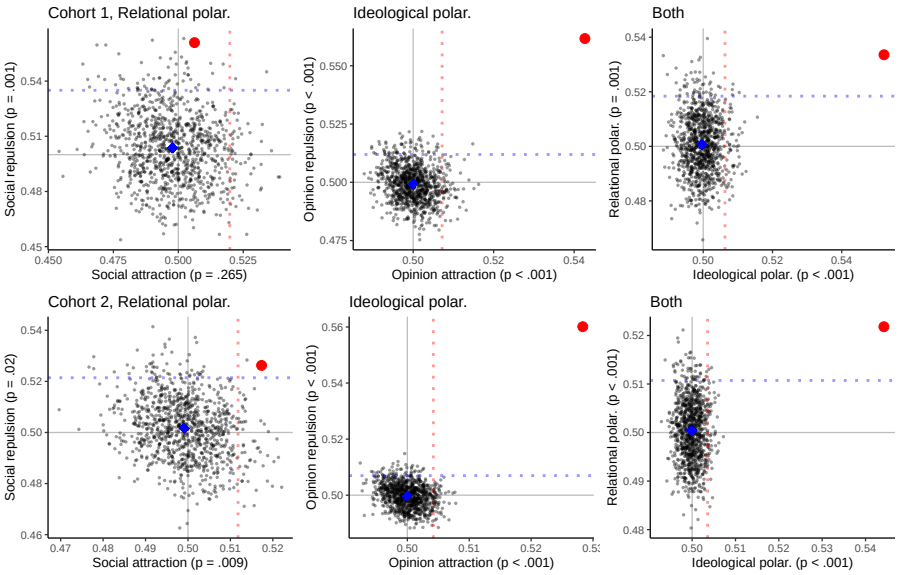
Wave 2


Cohort 1, Relational polar.
Ideological polar.
Both
Social repulsion (p = .003)
Opinion repulsion (p < .001)
Relational polar. (p = .013)
Social attraction (p = .614)
Opinion attraction (p < .001)
Ideological polar. (p < .001)
Cohort 2, Relational polar.
Ideological polar.
Both
Social repulsion (p = .105)
Opinion repulsion (p < .001)
Relational polar. (p = .113)
Social attraction (p = .45)
Opinion attraction (p < .001)
Ideological polar. (p < .001)

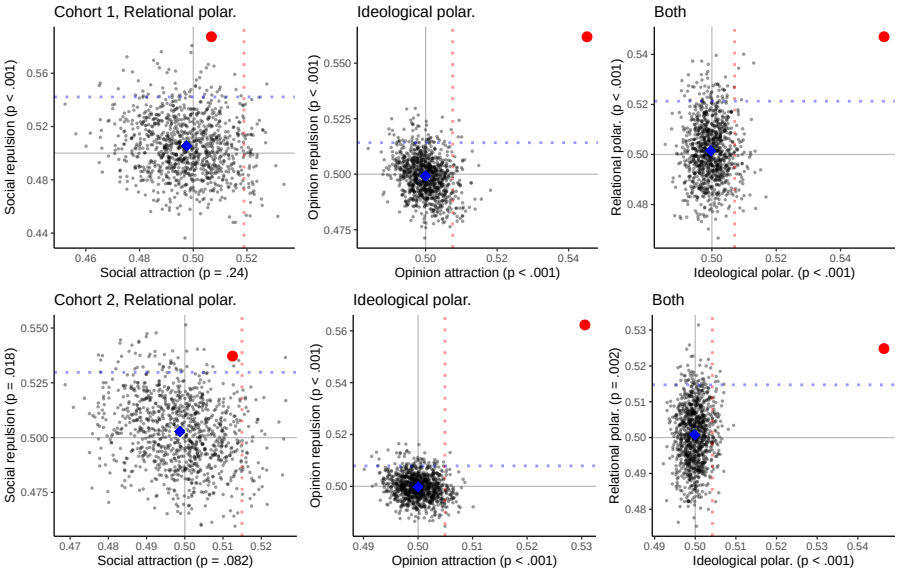
Wave 3


Cohort 1, Relational polar.
Ideological polar.
Both
Social repulsion (p = .187)
Opinion repulsion (p < .001)
Relational polar. (p = .246)
Social attraction (p = .607)
Opinion attraction (p < .001)
Ideological polar. (p < .001)
Cohort 2, Relational polar.
Ideological polar.
Both
Social repulsion (p = .079)
Opinion repulsion (p < .001)
Relational polar. (p = .007)
Social attraction (p = .073)
Opinion attraction (p < .001)
Ideological polar. (p < .001)

Wave 4


Cohort 1, Relational polar.
Ideological polar.
Both
Social repulsion (p = .001)
Opinion repulsion (p < .001)
Relational polar. (p = .001)
Social attraction (p = .265)
Opinion attraction (p < .001)
Ideological polar. (p < .001)
Cohort 2, Relational polar.
Ideological polar.
Both
Social repulsion (p = .02)
Opinion repulsion (p < .001)
Relational polar. (p < .001)
Social attraction (p = .009)
Opinion attraction (p < .001)
Ideological polar. (p < .001)

Wave 5


Cohort 1, Relational polar.
Ideological polar.
Both
Social repulsion (p < .001)
Opinion repulsion (p < .001)
Relational polar. (p < .001)
Social attraction (p = .24)
Opinion attraction (p < .001)
Ideological polar. (p < .001)
Cohort 2, Relational polar.
Ideological polar.
Both
Social repulsion (p = .018)
Opinion repulsion (p < .001)
Relational polar. (p = .002)
Social attraction (p = .082)
Opinion attraction (p < .001)
Ideological polar. (p < .001)

References

Block, P. (2015). Reciprocity, transitivity, and the mysterious three-cycle. *Social Networks*, 40:163–173.

Davis, J. A. (1970). Clustering and hierarchy in interpersonal relations: Testing two graph theoretical models on 742 sociomatrices. *American Sociological Review*, 35(5):843–851.

Holland, P. W. and Leinhardt, S. (1977). A method for detecting structure in sociometric data. In *Social networks*, pages 411–432. Elsevier.

Lazer, D., Rubineau, B., Chetkovich, C., Katz, N., and Neblo, M. (2010). The coevolution of networks and political attitudes. *Political Communication*, 27(3):248–274.

Ripley, R. M., Snijders, T. A., Boda, Z., Vörös, A., and Preciado, P. (2019). *Manual for SIENA version 4.0*. Oxford.

Vörös, A., Boda, Z., Raabe, I., Elmer, T., Mepham, K., Hoffman, M., and Stadtfeld, C. (2021). The swiss studentlife study: Investigating the emergence of an undergraduate community through dynamic, multidimensional social network data. *Social Networks*, 65:71–84.

Wang, C., Lizardo, O., and Hachen, D. S. (2020). Neither influence nor selection: Examining co-evolution of political orientation and social networks in the netsense and nethealth studies. *PLoS ONE*, 15(5):e0233458.