Online Appendix for "How Alliances Form and Conflict Ensues"

Appendix A. Equilibrium characterization

In this appendix, we provide the complete Nash equilibrium characterization for our 4-person game, with all equilibria summarized in Figure 1 in the main text. We consider the case of $c < k$ and the case of $k < c < 2k$ separately.

- $\boldsymbol{c < k}$

Depending on whether there is any negative link in equilibrium, all equilibria can be sorted into the following two categories, each with conditions that should be satisfied (and are easily verified) in equilibrium.

Category I (with no negative links): $\forall i, j, n_i(\boldsymbol{g}) = n_j(\boldsymbol{g})$.[1]

Category II (with at least one negative link): (1) $\forall i, j$ with $g_{i,j} = -1$, $n_i(\boldsymbol{g}) - n_j(\boldsymbol{g}) \geq 1$; (2) $\forall i, j$ with $n_i(\boldsymbol{g}) - n_j(\boldsymbol{g}) \geq 1$ and $\bar{g}_{i,j} = 1$, $\exists l \notin \{i, j\}$ such that $g_{i,l} = -1$.[2]

- **Category I (with no negative links): $\forall \boldsymbol{i, j, n_i(g) = n_j(g)}$.**

There are four possibilities: $\forall i, \exists z \in \{0,1,2,3\}, n_i(\boldsymbol{g}) = z$;

(1) $\forall i, n_i(\boldsymbol{g}) = 3$, that is equilibrium (2), that is, the Peace equilibrium. Uniqueness is straightforward.
(2) $\forall i, n_i(\boldsymbol{g}) = 2$, that is equilibrium (4). For uniqueness, WLOG, suppose player 1's friends are players 2 and 3. Then players 2 and 3 cannot be friends, otherwise player 4 will have no friends, violating $n_4(\boldsymbol{g}) = 2$. Thus, players 2 and 3 each will be friends with player 4, in which case player 4 also have 2 friends.
(3) $\forall i, n_i(\boldsymbol{g}) = 1$, that is equilibrium (6). For uniqueness, WLOG, suppose player 1's friend is player 2. Then player 3 cannot make friends with either player 1 or 2, otherwise player 1 or 2 will have more than one friend, violating $n_i(\boldsymbol{g}) = 1$. Thus, player 3 can only be friends with player 4, in which case player 4 also satisfies $n_1(\boldsymbol{g}) = 2$.
(4) $\forall i, n_i(\boldsymbol{g}) = 0$, that is equilibrium (8). Uniqueness is straightforward.

[1] When $c < k$, suppose there exist i, j such that $n_i(\boldsymbol{g}) - n_j(\boldsymbol{g}) \geq 1$, then player i has a profitable deviation by attacking player j, contradicting the equilibrium requirement with no negative links.

[2] Condition (1) is necessary since attacking a target who has no fewer friends is not profitable when $c < k$. Condition (2) is necessary since when $c < k$ the only reason that player i chooses not to attack player j who has fewer friends than himself is that he has another target l to attack. In that case, any profit player i would gain from attacking j will be completely offset by the lower payoff from attacking l since j would become i's enemy rather than a friend who could have helped i in the event of attacking l.

- **Category II (with at least a negative link): (1) $\forall i, j$ with $g_{i,j} = -1$, $n_i(g) - n_j(g) \geq 1$; (2) $\forall i, j$ with $n_i(g) - n_j(g) \geq 1$ and $\bar{g}_{i,j} = 1$, $\exists l \notin \{i, j\}$ such that $g_{i,l} = -1$.**

WLOG, suppose $g_{1,4} = -1$, that is player 1 attacks player 4. Then we must have $n_1(\boldsymbol{g}) - n_4(\boldsymbol{g}) \geq 1$. This implies either scenario (1) player 1 makes friends with both players 2 and 3 ($n_1(\boldsymbol{g}) = 2$), and player 4 must have at most one friend ($n_4(\boldsymbol{g}) \leq 1$), or scenario (2) player 1 makes friends with either player 2 or 3 ($n_1(\boldsymbol{g}) = 1$), and player 4 must have no friends ($n_4(\boldsymbol{g}) = 0$).

For scenario (1), there are three possibilities.

(i) Player 4 has no friends, and players 2 and 3 are friends with each other. In this case, it is easy to verify that there is a unique equilibrium with positive links being $\bar{g}_{1,2} = \bar{g}_{1,3} = \bar{g}_{2,3} = 1$ and negative links being $g_{1,4} = g_{2,4} = g_{3,4} = -1$. This corresponds to equilibrium (1), the Bully equilibrium.

(ii) Player 4 has no friends, and players 2 and 3 are not friends with each other. In this case, both players 2 and 3 also have incentive to attack player 4. It is easy to verify that there is a unique equilibrium with the positive links being $\bar{g}_{1,2} = \bar{g}_{1,3} = 1$ and the negative links being $g_{1,4} = g_{2,4} = g_{3,4} = -1$, resulting in $2 = n_1(\boldsymbol{g}) > n_2(\boldsymbol{g}) = n_3(\boldsymbol{g}) = 1 > n_4(\boldsymbol{g}) = 0$. This is equilibrium (9).

(iii) Player 4 has one friend, and WLOG suppose player 4's only friend is player 2. In this case, players 1 and 2 each have two friends, and players 3 and 4 each have only one friend. Thus, player 2 will have incentive to attack player 3. It is easy to verify that there is a unique equilibrium with the positive links being $\bar{g}_{1,2} = \bar{g}_{1,3} = \bar{g}_{2,3} = 1$ and the negative links being $g_{1,4} = g_{2,3} = -1$, resulting in $2 = n_1(\boldsymbol{g}) = n_2(\boldsymbol{g}) > n_3(\boldsymbol{g}) = n_4(\boldsymbol{g}) = 1$. This is equilibrium (11).

Next, we consider scenario (2). WLOG suppose player 1's only friend is player 2. Since player 4 has no friend and player 1 is not player 3's friend, player 3 can at most make friends with 2. There are two possibilities.

(i) Player 3 has no friends. In this case, players 1 and 2, who are friends with each other, have incentive to attack both players 3 and 4, who have no friends. Thus, there is a unique equilibrium in this case with the only positive link being $\bar{g}_{1,2} = 1$ and the negative links being $g_{1,3} = g_{1,4} = g_{2,3} = g_{2,4} = -1$, resulting in $1 = n_1(\boldsymbol{g}) = n_2(\boldsymbol{g}) > n_3(\boldsymbol{g}) = n_4(\boldsymbol{g}) = 0$. This is equilibrium (12).

(ii) Player 3 has one friend, which is player 2. In this case, both players 1 and 3 are player 2's friends, and both players 2 and 3 also have incentive to attack player 4. It is easy to verify that there is a unique equilibrium with the positive links being $\bar{g}_{1,2} = \bar{g}_{2,3} = 1$ and the negative links being $g_{1,4} = g_{2,4} = g_{3,4} = -1$, resulting in $2 = n_2(\boldsymbol{g}) > n_1(\boldsymbol{g}) = n_3(\boldsymbol{g}) = 1 > n_4(\boldsymbol{g}) = 0$. By relabeling the players, this is essentially equilibrium (9).

- $\boldsymbol{k < c < 2k}$

Depending on whether there is any negative link in equilibrium, all equilibria can be sorted into two categories, each with conditions that should be satisfied (and are easily verified) in equilibrium.

Category I (with no negative links): $\forall i, j, \left|n_i(\boldsymbol{g}) - n_j(\boldsymbol{g})\right| \leq 1$.[3]

Category II (with at least a negative link): $\forall i, j$, $g_{i,j} = -1$ if and only if $n_i(\boldsymbol{g}) - n_j(\boldsymbol{g}) \geq 2$.[4]

- **Category I (with no negative links): $\forall \boldsymbol{i}, \boldsymbol{j}, \left|\boldsymbol{n_i}(\boldsymbol{g}) - \boldsymbol{n_j}(\boldsymbol{g})\right| \leq \mathbf{1}$.**

Analogous to the Category I analysis for the case of $c < k$, it is straightforward to show that all networks such that $\forall i, j, n_i(\boldsymbol{g}) = n_j(\boldsymbol{g})$ are equilibria, which include (2), (4), (6) and (8).

Since $\forall i, j, \left|n_i(\boldsymbol{g}) - n_j(\boldsymbol{g})\right| \leq 1$, for equilibria in which not all players have the same number of friends, there are in total three possible scenarios.

(1) WOLG, suppose $3 = n_1(\boldsymbol{g}) \geq max\{n_2(\boldsymbol{g}), n_3(\boldsymbol{g})\} \geq min\{n_2(\boldsymbol{g}), n_3(\boldsymbol{g})\} \geq n_4(\boldsymbol{g}) = 2$. Uniqueness: $n_1(\boldsymbol{g}) = 3$ means players 2, 3, and 4 are all player 1's friends. WLOG, suppose the other friend of player 4 is player 2. Note that player 3 should have at least one additional friend other than player 1, since $n_3(\boldsymbol{g}) \geq 2$; however, player 3's friend cannot be player 4 since player 4 already has 2 friends (players 1 and 2). This implies that player 3 must be friends with player 2. Thus, there is a unique equilibrium in this scenario with the only zero link being $\bar{g}_{3,4} = 0$, resulting in $3 = n_1(\boldsymbol{g}) = n_2(\boldsymbol{g}) > n_3(\boldsymbol{g}) = n_4(\boldsymbol{g}) = 2$. This is equilibrium (3).

(2) WOLG, suppose $2 = n_1(\boldsymbol{g}) \geq max\{n_2(\boldsymbol{g}), n_3(\boldsymbol{g})\} \geq min\{n_2(\boldsymbol{g}), n_3(\boldsymbol{g})\} \geq n_4(\boldsymbol{g}) = 1$. There are two cases.

   (i) Suppose player 4 is player 1's friend, and WLOG, suppose the other friend of player 1 is player 2. Note that player 3 cannot be friends with either player 1 or 4, since player 1 already has 2 friends and player 4 already has 1 friend. Thus, player 3 must be friends with player 2, since player 3 should have at least one friend. Thus, there is a unique equilibrium with the only positive links being $\bar{g}_{1,2} = \bar{g}_{1,4} = \bar{g}_{2,3} = 1$, resulting in $2 = n_1(\boldsymbol{g}) = n_2(\boldsymbol{g}) > n_3(\boldsymbol{g}) = n_4(\boldsymbol{g}) = 1$.

   (ii) Suppose player 1's friends are players 2 and 3. Note that player 4 must have exactly one friend, who cannot be player 1; WLOG, let player 4's friend be player 2. Now player 2 already has two friends and cannot have any additional friend. Since players 1, 2, and 4 are all capped with the maximum number of friends they can have, player 3 cannot have any additional friend. This is the equilibrium with the only positive links being $\bar{g}_{1,2} = \bar{g}_{1,3} = \bar{g}_{2,4} = 1$, resulting in $2 = n_1(\boldsymbol{g}) = n_2(\boldsymbol{g}) > n_3(\boldsymbol{g}) = n_4(\boldsymbol{g}) = 1$.

---

[3] When $k < c < 2k$, suppose there exist i, j such that $n_i(\boldsymbol{g}) - n_j(\boldsymbol{g}) \geq 2$, then player i has a profitable deviation by attacking player j, contradicting the equilibrium requirement with no negative links.

[4] The "only if" part is straightforward since $k < c < 2k$. To see that the "if" part holds, note that the only possible exception would be that player i chooses not to attack player j because he attacks another player l as in the case of $c < k$. In that situation, however, player i must have exactly two friends (he cannot befriend player l) and one of them must be player j. But this contradicts the condition that player i has at least two more friends than player j.

By relabeling, these two cases lead to the same equilibria, represented by equilibrium (5).

(3) WOLG, suppose $1 = n_1(\boldsymbol{g}) \geq max\{n_2(\boldsymbol{g}), n_3(\boldsymbol{g})\} \geq min\{n_2(\boldsymbol{g}), n_3(\boldsymbol{g})\} \geq n_4(\boldsymbol{g}) = 0$. Uniqueness: WLOG, suppose player 1's friend is player 2. Then player 2 cannot have any additional friend. Also note that player 3 cannot have any friend, since players 1 and 2 are friends with each other and player 4 has no friends. Thus, there is a unique equilibrium in this scenario with the only positive link being $\bar{g}_{1,2} = 1$, resulting in $1 = n_1(\boldsymbol{g}) = n_2(\boldsymbol{g}) > n_3(\boldsymbol{g}) = n_4(\boldsymbol{g}) = 0$. This is equilibrium (7).

- **Category II (with at least a negative link): $\forall \boldsymbol{i, j}$, $\boldsymbol{g_{i,j} = -1}$ if and only if $\boldsymbol{n_i(g) - n_j(g) \geq 2}$.**

WLOG, suppose $g_{1,4} = -1$, that is player 1 attacks player 4. We have $n_1(\boldsymbol{g}) - n_4(\boldsymbol{g}) \geq 2$. This implies that the following two conditions hold: (i) player 1 must be friends with players 2 and 3, $n_1(\boldsymbol{g}) = 2$; and (ii) player 4 must have no friends, $n_4(\boldsymbol{g}) = 0$.

There are two possibilities.

(1) Suppose players 2 and 3 are friends with each other. In this scenario, we have $n_2(\boldsymbol{g}) = n_3(\boldsymbol{g}) = 2$. Thus, both players 2 and 3 will also attack player 4, resulting in an equilibrium with positive links being $\bar{g}_{1,2} = \bar{g}_{1,3} = \bar{g}_{2,3} = 1$ and negative links being $g_{1,4} = g_{2,4} = g_{3,4} = -1$. That is equilibrium (1), the Bully equilibrium.

(2) Suppose players 2 and 3 are not friends with each other. In this scenario, we have $n_2(\boldsymbol{g}) = n_3(\boldsymbol{g}) = 1$. Thus, neither player 2 nor 3 will attack player 4, resulting in an equilibrium with positive links being $\bar{g}_{1,2} = \bar{g}_{1,3} = 1$ and a negative link being $g_{1,4} = -1$. This is equilibrium (10).

Appendix B. Experimental instructions (English translation)

General Information:

You are participating in a decision-making study. Please read the following instructions carefully. These instructions are the same for all the participants. During the experiment, you are not allowed to communicate with other participants. Turn-off your mobile phone and put it in the envelope on your desk. If you have any questions, please raise your hand. One of the experimenters will approach you to answer your question.

You have earned 15 RMB for showing up on time. You can earn additional money by means of earning points during the experiment. The number of points that you earn depends on your own choices and the choices of other participants. At the end of the experiment, the total number of points that you earn during the experiment will be exchanged at the rate of:

5 points = 1 RMB

The money you earn will be paid out in cash via WeChat. Your decisions in this experiment will be anonymous, meaning no one can associate your name with your action throughout this study, and no other participants will be able to see how much you earn.

Overview of the experiment:

The experiment consists of 5 blocks, each of which has 4 rounds. There are 20 rounds in total. At the beginning of each round, you will be randomly matched with three other participants. Each participant's position in your group will be shown as a circle on a specific position of the screen (either upper, lower, left or right, see screenshot below). The green circle represents yourself, while the black circles represent other three participants in your group. These participants are all currently in this room, but everyone's identity will be anonymous. The groups and the positions within a group will change across rounds.

Your decisions:

During each round, you may connect to one or more of the other participants in your group via two different means (you can also choose not to connect):

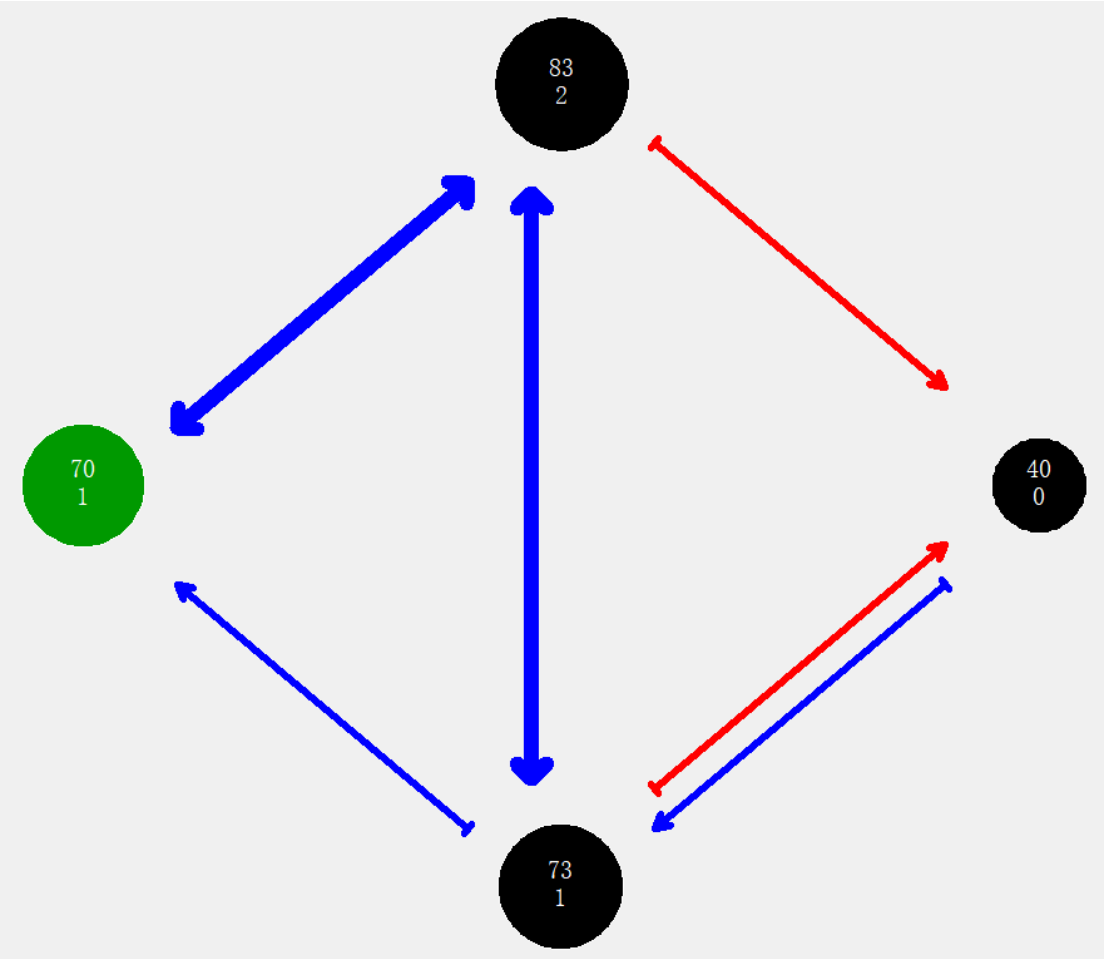
1. If you LEFT-MOUSE-CLICK on one of the black circles representing another participant, a blue link with an arrow pointing to that participant will appear. Left-clicking again on that participant, the blue link will be removed.

- Initiating a blue link represents an attempt to establish a partnership with another participant. Importantly, a partnership is only effective if ***both*** participants have initiated blue links, otherwise the partnership is ineffective. On the screen, if both participants initiate a blue link to each other, the blue link then becomes a bold double-headed arrow link (see screenshot below, the upper & lower players, and the upper & left players). Only in this case, the partnership is effective. Note that as long as one of the participants removes the link, the partnership will be ineffective.

2. If you RIGHT-MOUSE-CLICK on another participant, a red link with an arrow pointing to that participant will appear. Right-clicking again on that participant, the red link will be removed.

- Initiating a red link represents establishing a competitive relationship with another participant. Any unilateral initiation of a red link is effective (see screenshot below, both the upper and lower players initiate a red link to the right player). It means that a competitive relationship is effective as long as ***at least one side*** initiates a red link.

When you are making your linking decisions, you will be able to see your group members making and removing links simultaneously in real time. Likewise, other members in your group can see your making and removing decisions simultaneously in real time.



The number on top of the green circle indicates your current points, and the number on top of the black circle indicates other's current points. The size of a circle changes with the points that a player will receive: a larger circle means that that participant receives more points. The bottom

number in each circle indicates the number of effective blue (partnership) links a player currently has.

Remarks:

- Note that to change a blue link to a red link, or vice versa, there is no need to 'unlink' the previous choice. You can simply directly left-click for blue or right-click for red. You cannot initiate a blue link and a red link to the same other participant at the same time.
- Note that there may be a slight time-lag between your click and the changes of the numbers on the screen. One click is enough to change a link successfully. A subsequent click will not be effective until the previous click is successfully in place. Therefore, be patient until a link is changed in order to make subsequent changes.

Your earnings:

Below we explain how to calculate your points for each round. Points depend on the links you and other participants make. Read this carefully. Do not worry if you find it difficult to grasp immediately—recall that the concurrent point values will be shown as the top number in each circle representing a player. We present an example with calculations below.

At the beginning of each round, each of the players will receive an endowment of 70 points. Note you start with 70 points every round. Formation of blue links is costless, while initiating each red link costs some points, which are either 3, 5, 7, 9 and 11. Before a block begins, you will know the cost of a red link for the 4 rounds in this block. You also know that the cost is the same for all members in your group.

If a player neither initiates nor receives a red link, her points remain as 70.

In the presence of a red link between say player A and player B, the point change depends on the difference between A's effective blue (partnership) links and B's effective blue (partnership) links. Take player A as an example, if A initiated a red link to B, A's additional points from the competitive relationship with player B are:

10*(A's effective blue links – B's effective blue links) – costs of red links

If A did not initiate a red link (thus it must be B who initiated a red link to A), A's additional points from the competitive relationship with player B are:

10*(A's effective blue links – B's effective blue links)

Point changes are calculated separately for each red link that you initiate or receive. Therefore, the total points, which are shown as the top number in each circle, are the sum of the endowment and point changes across all of your existing red links.

In the example shown in the above figure, the cost of each red link is 7 points:

The upper player

- initiates a red link to the right player;

- has two effective blue links with the lower and left players respectively, while the right player has no effective blue link;
- has payoff = 70 + 10(2-0) – 7 = 83.

The right player

- receives two red links from the upper and lower players respectively;
- has no effective blue link, while the upper player has two effective blue links and the lower player has one effective blue link;
- has payoff = 70 + 10(0-2) + 10(0-1) = 40.

The lower player

- initiates a red link to the left player;
- has one effective blue link, while the left player has no effective blue link;
- has payoff = 70 + 10*(1-0) – 7 = 73.

The left player

- neither initiates nor receives a red link;
- has one effective blue link;
- has payoff = 70.

Each player's final points in that round are determined at the end of that round. You can make as many adjustments of links as you like during a round; these adjustments are free. Both links and points in the circles are updated in real time. However, once that round ends, your points are determined by whatever the situation is in terms of your links at that point in time. Each round lasts somewhere between 75 and 105 seconds. The end will be at an unknown and random moment within this time interval. Please note that different rounds will not last equally long.

The computer will randomly choose one block (4 rounds) to calculate the total points as your final earnings. To give yourself the best chance of earning the most, you should decide carefully about every single round.

Questionnaire:

After the 20 rounds, you will be asked to fill in a brief questionnaire. Please take your time to fill in this questionnaire accurately. After you finish the questionnaire, the total amount you have earned from this experiment will be shown on the computer screen. Please remain seated until being instructed to leave.

This concludes the instructions. To make sure that everyone understands the instructions, you will now be asked to answer some comprehension questions. Please raise your hand if you need help. We will start the experiment once every participant has correctly answered all the comprehension questions.

Comprehension Quiz:

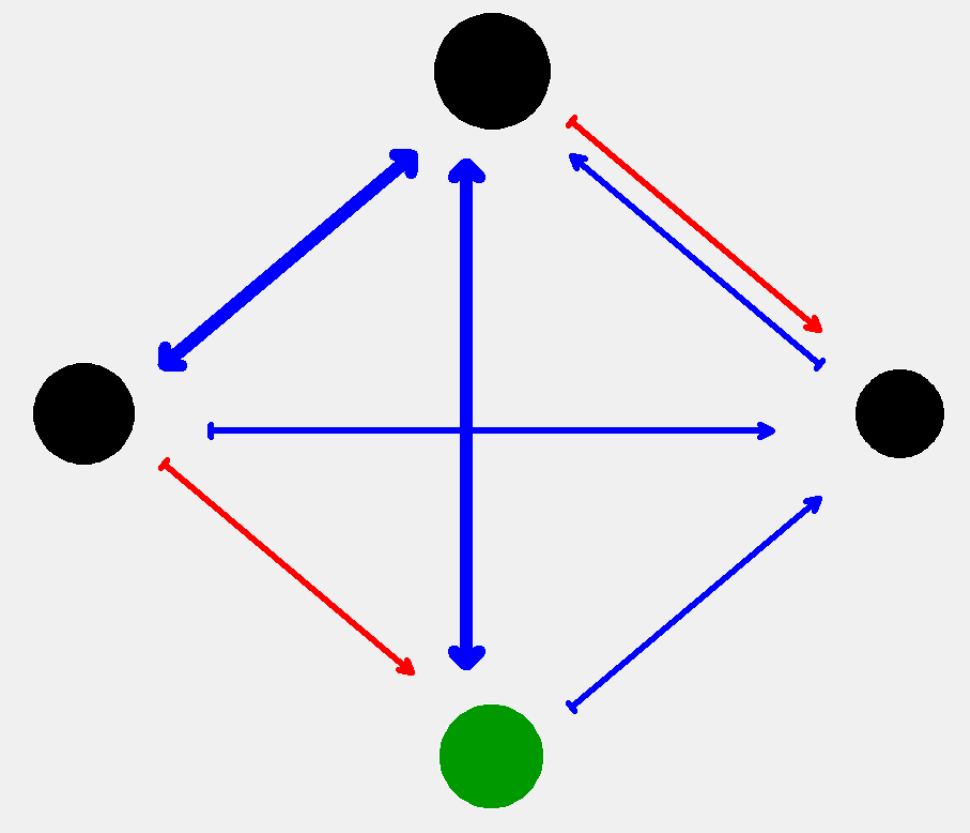
(on participants' computer screens)

General quiz (True or False):

- Your actions are anonymous in this study.
- You will receive money in cash via WeChat.
- In each block, you meet the same three other participants, but you don't know who they are. And their positions on the screen are unchanged within a block.
- You will meet the same three other participants from block to block.
- Every participant gets 70 points and starts with 70 points at the beginning of each round.
- Each round will end between 75 and 105 seconds, and the ending time for each round is different.
- If a round ends at the 105-th second, then your points in that round are determined by the nature of all links at the 105-th second.

Quiz on calculating the payoff:

Suppose a round ends at the 120th second, and at that moment, players have the following link configuration:



Upper player:

- Has initiated __ red links and received __ red links.
- Has __ effective blue links.
- The final payoff is __ points.

Bottom player:

- Has initiated __ red links and received __ red links.
- Has __ effective blue links.
- The final payoff is __ points.

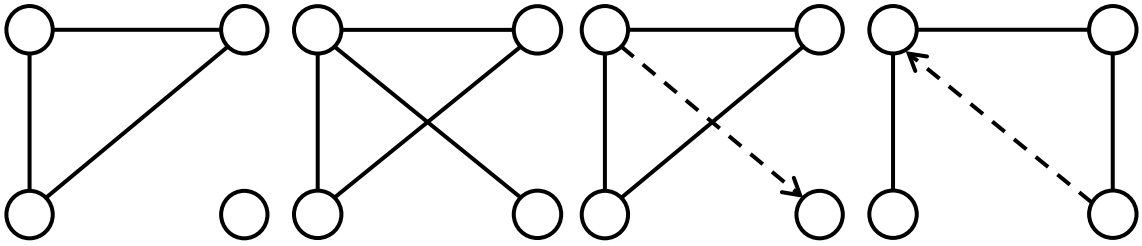
Left player:

- Has initiated __ red links and received __ red links.
- Has __ effective blue links.
- The final payoff is __ points.

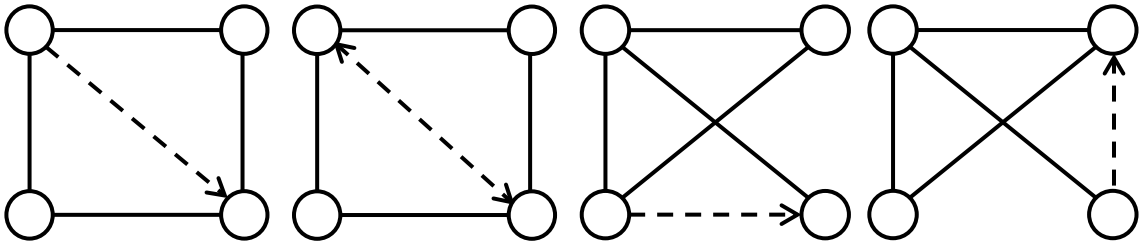
Right player:

- Has initiated __ red links and received __ red links.
- Has __ effective blue links.
- The final payoff is __ points.

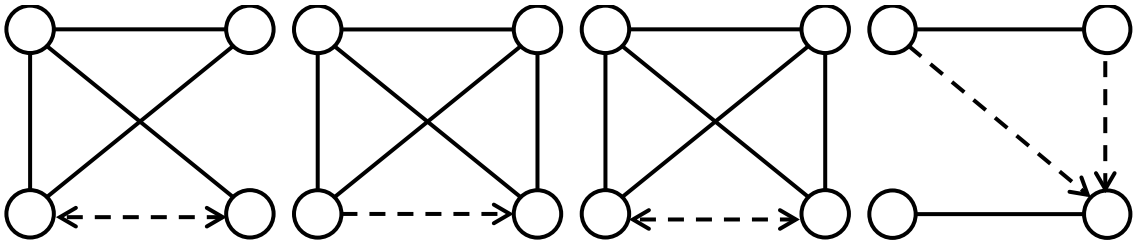
Appendix C. Additional figures and tables



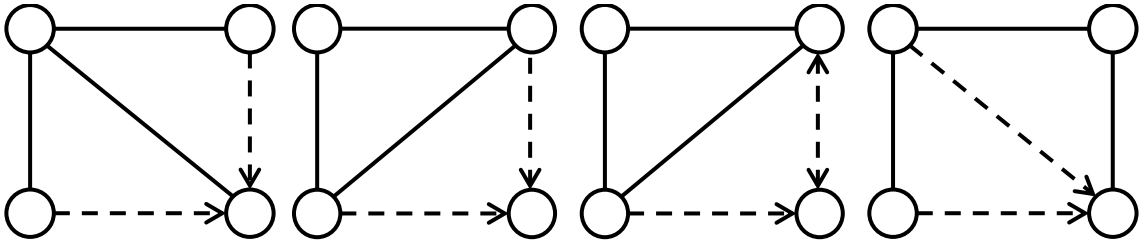
(13) (14) (15) (16)



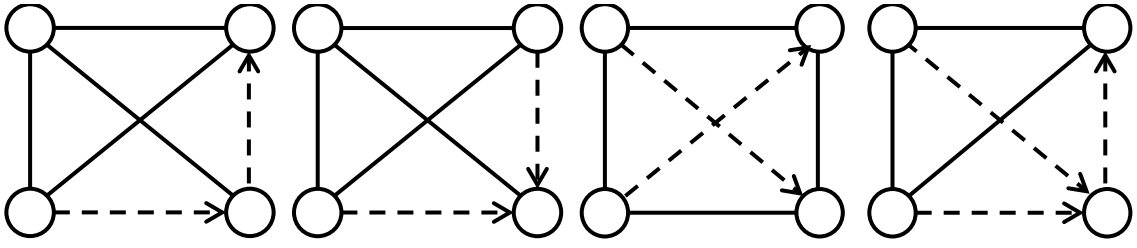
(17) (18) (19) (20)



(21) (22) (23) (24)



(25) (26) (27) (28)



(29) (30) (31) (32)

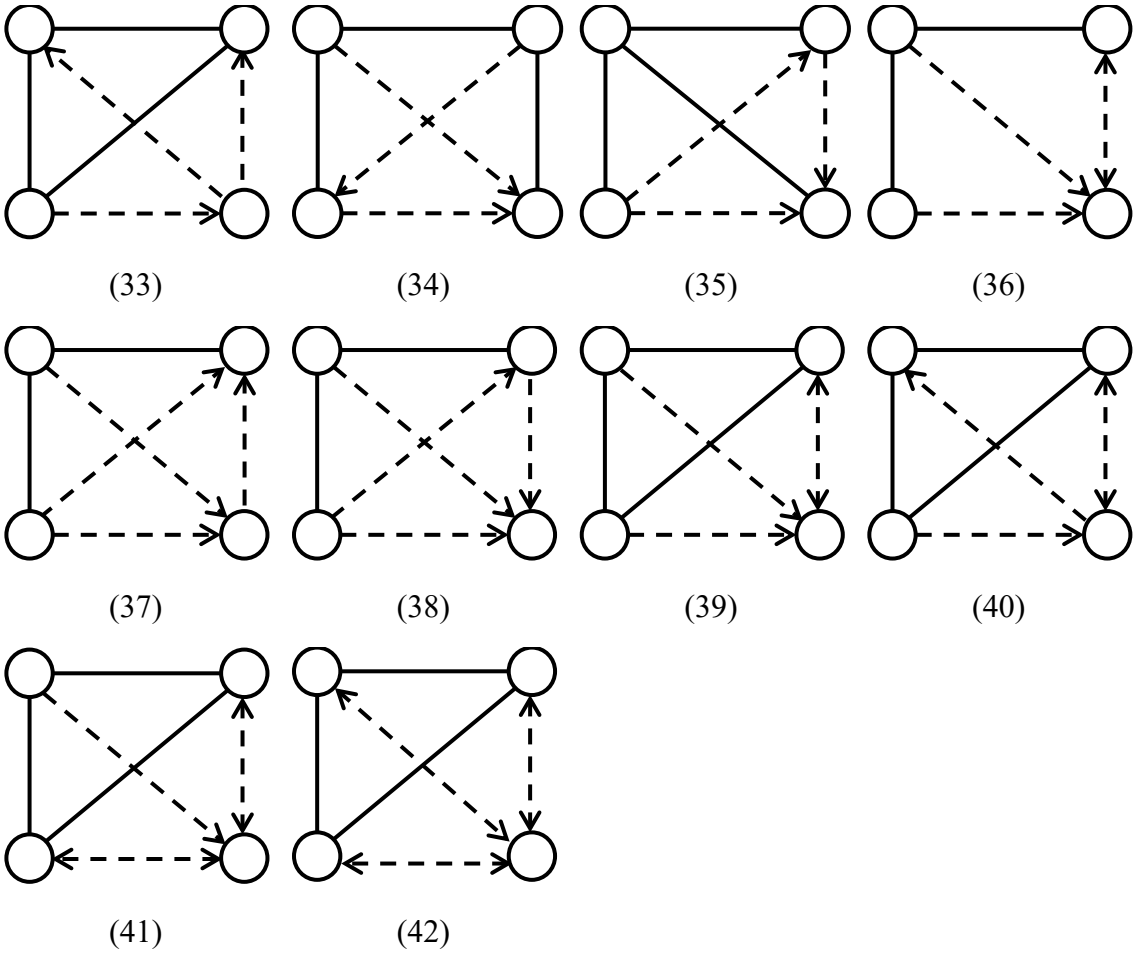


Figure C1. Non-equilibrium final network types formed at least once

Table C1. Frequency of final network types

	Prominent equilibrium type		Equilibrium types (3)-(12)		Non-equilibrium types (13)-(42)		
	Bully (1)	Peace (2)	(3)	Other (4)-(12)	(30)	(39)	Other
Within-Subject Experiment							
Cost = 3	57.9	23.8	1.8	0.6	7.9	4.3	3.7
Cost = 5	59.1	29.9	0	1.2	3.7	1.8	4.3
Cost = 7	52.4	37.2	1.2	0	7.3	0.6	1.2
Cost = 9	45.1	44.5	1.2	1.2	4.3	1.8	1.8
Cost = 11	34.1	59.1	3.0	0	0.6	0	3.0
Between-Subject Experiment							
Cost = 3	67.2	20.6	1.3	1.3	3.1	2.8	3.8
Cost = 7	46.6	44.7	1.9	0	3.4	0.9	2.5
Cost = 11	22.2	65.3	5.9	0.3	0.3	0	5.9

Note: Please refer to Figure 4 and Figure C1 for all network types and their associated ID number.

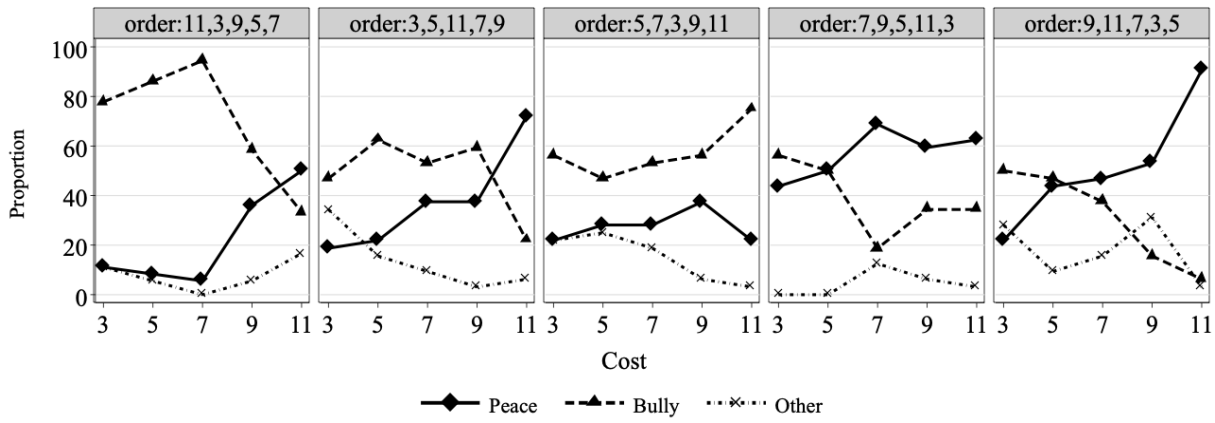


Figure C2. Frequency of final network types for each ordering of treatments

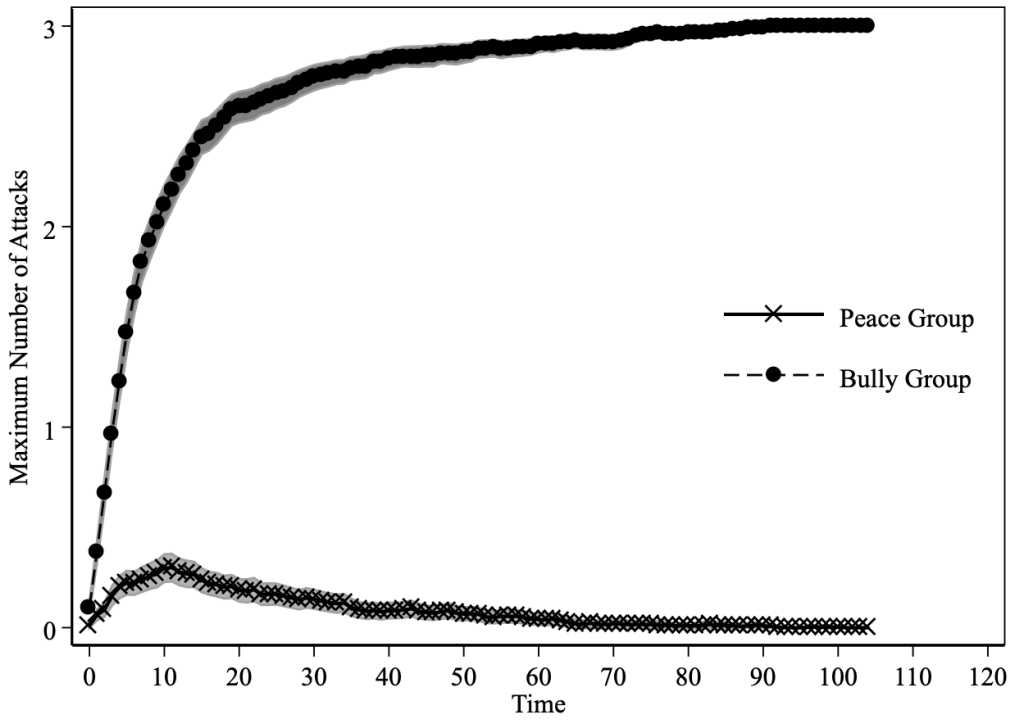


Figure C3: Evolution of maximum number of attacks received by any player per group – within-subject experiment

Note: This Figure shows the maximum number of attacks received by any player in a group. By definition, no player receives attacks at the end in Peace groups, and a player (final victim) receives 3 attacks at the end in Bully groups. The grey shaded area indicates 95% confidence intervals.

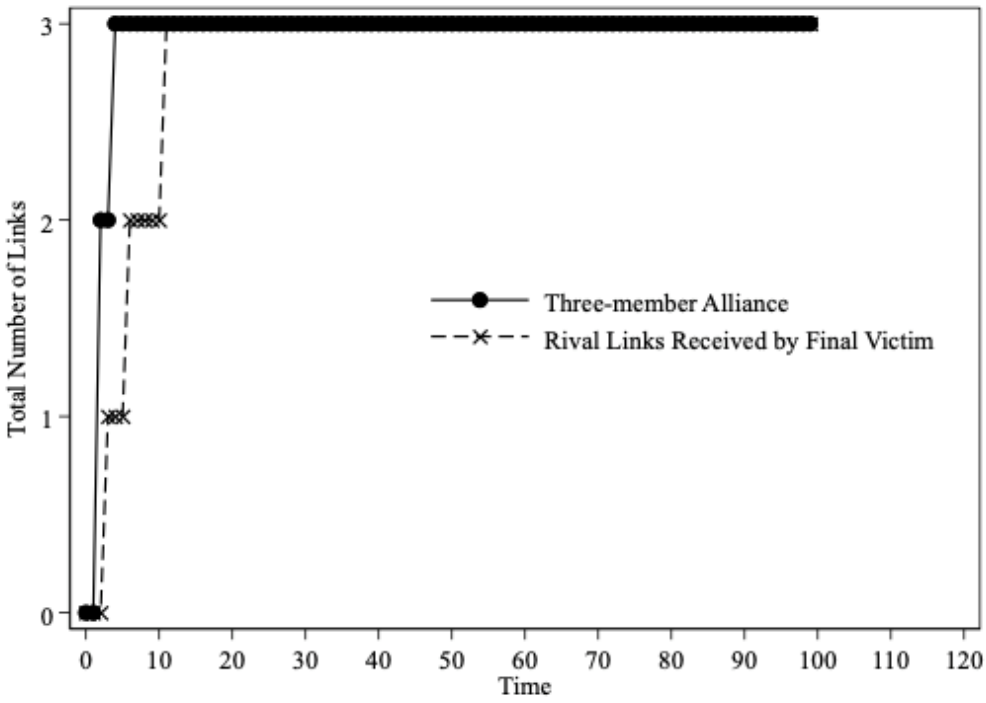


Figure C4: The evolution of median attacks received by the final victim and median effective friendships among the other three players in Bully groups – within-subject experiment

Appendix D. Dynamics of network formation separately for different treatments/cost levels – Within-subject experiment

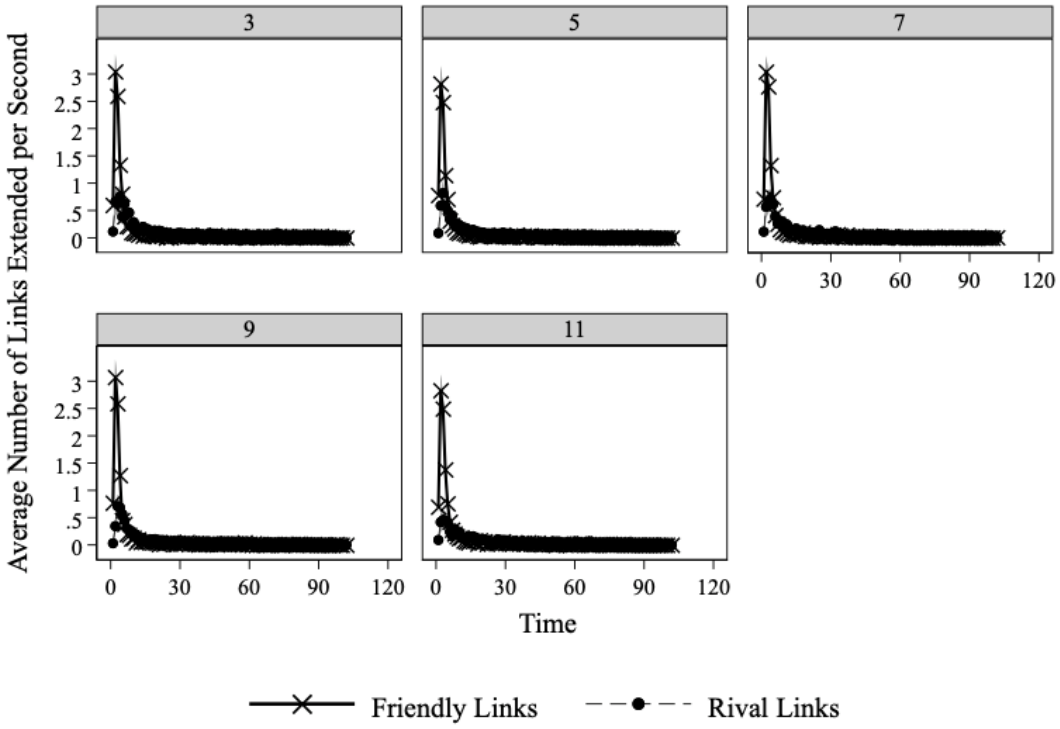


Figure D1: Extension of links per group per second by cost level

Note: The grey shaded area indicates 95% confidence intervals.

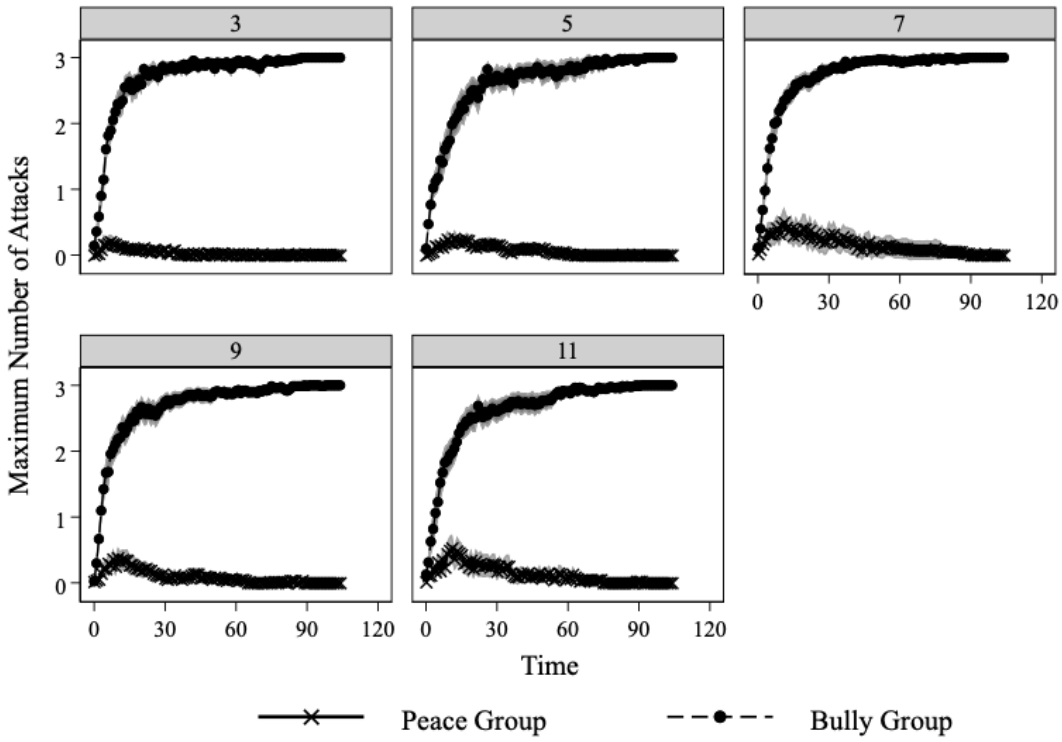


Figure D2: The evolution of the maximum number of attacks received by any player per group by cost level

Note: The grey shaded area indicates 95% confidence intervals.

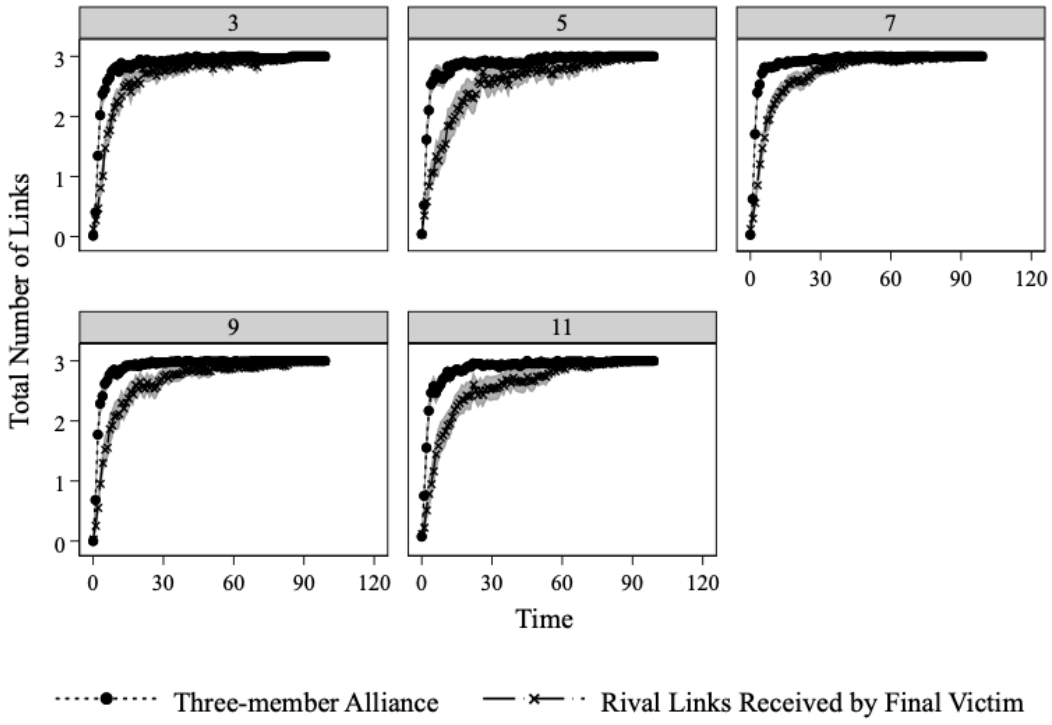


Figure D3: The evolution of average attacks received by the final victim and average effective friendships among the other three players in Bully groups by cost level

Note: The grey shaded area indicates 95% confidence intervals.

Table D1: The pattern of transition to final victims by cost level

Type	N	% Receive 1 attack	% Receive 2 attacks	% Receive 3 attacks	% Final victim
Cost=3					
First victim	140	100%	70.7%	55.0%	52.9%
Initiator	140	63.6%	22.9%	13.6%	12.9%
Others	376	24.2%	5.9%	4.0%	2.1%
Cost=5					
First victim	132	100%	65.9%	52.3%	50.0%
Initiator	132	67.4%	27.3%	22.0%	18.9%
Others	392	18.4%	4.3%	3.1%	1.0%
Cost=7					
First victim	132	100%	70.5%	55.3%	51.5%
Initiator	132	60.6%	14.4%	11.4%	10.6%
Others	392	16.8%	2.8%	1.8%	0.8%
Cost=9					
First victim	120	100%	64.2%	51.7%	50.0%
Initiator	120	54.2%	12.5%	9.2%	9.2%
Others	416	14.7%	2.6%	1.4%	0.5%
Cost=11					
First victim	106	100%	58.5%	42.5%	36.8%
Initiator	106	50.9%	19.8%	12.3%	11.3%
Others	444	13.3%	2.3%	1.1%	0.7%

Appendix E. The relationship between network patterns in the first few seconds and the final outcome

Given our findings of early divergence between Peace and Bully groups discussed in the previous subsection, we now consider whether particular categories of network formations are predictive of eventual convergence to Peace or Bully outcomes. To provide statistical evidence on factors that can explain the divergent paths of Bully and Peace networks, we turn to a group-level regression analysis with a binary dependent variable of whether a group eventually converges to Bully or Peace networks.

We define five key explanatory variables for this analysis. The first two binary variables relate to the pattern of forming alliance: *OneAlliance* indicates whether there is one and only one three-member alliance; *FullConnect* indicates whether all group members are mutual friends. The justification for these variables is that the formation of an alliance that is exclusive to the fourth member is might be a precondition to Bully networks, whereas the formation of four fully connected members is conducive to Peace networks. We thus hypothesize that *OneAlliance* predicts whether a group converges to Bully networks whereas *FullConnect* predicts whether a group reaches Peace networks.

The next two binary variables are related to the pattern of making rivals. *maxAttack1* indicates whether the maximum number of rival links received by any player in a group is equal to 1; and *maxAttack2* indicates whether the maximum number of rival links received by any player in a group is equal to 2. Since both variables measure different degrees of progress in coordinating on a common rival, we hypothesize that both *maxAttack1* and *maxAttack2* are predictive of Bully networks while *maxAttack2* has a stronger impact than *maxAttack1*.

We test how these state variables of the network status at time t (3~10 seconds) predict the final network, second by second. Table E1 reports the mean for each of these explanatory variables at each second from the third to the tenth second (the variables at the first two seconds are not included because there are very few observations. For ease of interpretation, we first consider only variables related to the pattern of alliance, then separately consider only variables related to the pattern of making rivals, and finally consider all variables together to see the relative importance of these two subsets of variables. The dependent variable for all Probit regressions below is whether the final network is Bully (as opposed to Peace).

Table E1: The means of all explanatory variables at each second

	3 sec	4 sec	5 sec	6 sec	7 sec	8 sec	9 sec	10 sec
OneAlliance	0.185	0.295	0.353	0.364	0.380	0.401	0.418	0.428
TwoAlliance	0.091	0.181	0.185	0.171	0.177	0.159	0.131	0.128
FullConnect	0.018	0.101	0.185	0.243	0.272	0.291	0.313	0.321
maxAttack1	0.252	0.268	0.240	0.207	0.204	0.164	0.150	0.138
maxAttack2	0.052	0.121	0.156	0.190	0.171	0.203	0.200	0.188
N	725	725	725	725	725	725	725	725

Table E2 reports estimates from regressions that include *OneAlliance* and *FullConnect*. It is striking how quickly the final outcome is resolved: *FullConnect* starts to negatively predict Bully networks by the 4th second, and continues to do so with generally increasing strength as the seconds proceed. For *OneAlliance*, the significant positive prediction of Bully networks starts in the 5th second, and generally strengthens as the seconds proceed, mostly up to the 8th second.

Table E2: Probit model estimates of network state variables (alliance variables)

	3 sec	4 sec	5 sec	6 sec	7 sec	8 sec	9 sec	10 sec
OneAlliance	-0.013	0.075*	0.107***	0.114***	0.173***	0.218***	0.202***	0.227***
	(0.064)	(0.043)	(0.035)	(0.017)	(0.035)	(0.037)	(0.034)	(0.028)
FullConnect	0.097	-0.295***	-0.388***	-0.432***	-0.423***	-0.381***	-0.384***	-0.354***
	(0.194)	(0.111)	(0.052)	(0.043)	(0.016)	(0.028)	(0.028)	(0.023)
N	725	725	725	725	725	725	725	725

Note: The dependent variable is 1 if the final status of the group is Bully, and 0 if Peace. We only include groups in which their final status is either Bully or Peace (725 out of 820 groups). The table reports average marginal effect estimates with standard errors clustered at the session level. All regressions include period fixed effects. * $p<0.1$; ** $p<0.05$; *** $p<0.01$

Table E3 reports estimates from regressions that include *maxAttack1* and *maxAttack2*. In general, 1 or 2 maximum attacks predicts Bully networks from the very beginning, and for the case of 1 maximum attack, the predictive power is not necessarily increasing in strength over time, while for 2 maximum attacks, the estimate is stable, and more influential than 1 maximum attack as the seconds go by.

Table E3: Probit model estimates of network state variables (attacking variables)

	3 sec	4 sec	5 sec	6 sec	7 sec	8 sec	9 sec	10 sec
maxAttack1	0.432***	0.394***	0.336***	0.309***	0.290***	0.225***	0.183**	0.191**
	(0.060)	(0.053)	(0.050)	(0.056)	(0.062)	(0.072)	(0.075)	(0.079)
maxAttack2	0.495***	0.524***	0.546***	0.520***	0.498***	0.496***	0.482***	0.413***
	(0.118)	(0.067)	(0.063)	(0.062)	(0.079)	(0.082)	(0.090)	(0.104)
N	725	725	725	725	725	725	725	725

Note: The dependent variable is 1 if the final status of the group is Bully, and 0 if Peace. The table reports average marginal effect estimates with standard errors clustered at the session level. All regressions include period fixed effects. * $p<0.1$; ** $p<0.05$; *** $p<0.01$

For completeness, we also include another variable, named *maxAttack3*, meaning that the maximum number of rival links received by any player in a group is 3. Thus, there is already a common rival in the group if *maxAttack3* = 1. This only happens starting from the fifth second. By definition, this variable is strongly correlated with *OneAlliance* as these two state variables often imply the realization of the Bully networks. Table E4 reports Probit estimates by including all three state variables related to attacking. Not surprisingly, *maxAttack3* strongly predicts Bully networks and its strength also tends to be the largest.

Table E4: Probit estimates of network state variables (all three attacking variables)

	5 sec	6 sec	7 sec	8 sec	9 sec	10 sec
maxAttack1	0.357***	0.338***	0.334***	0.300***	0.280***	0.292***
	(0.037)	(0.029)	(0.018)	(0.025)	(0.026)	(0.025)
maxAttack2	0.550***	0.520***	0.495***	0.490***	0.470***	0.427***
	(0.039)	(0.023)	(0.022)	(0.024)	(0.020)	(0.022)
maxAttack3	0.649***	0.684***	0.759***	0.689***	0.692***	0.690***
	(0.075)	(0.074)	(0.089)	(0.066)	(0.058)	(0.056)
N	725	725	725	725	725	725

Note: The dependent variable is 1 if the final status of the group is Bully, and 0 if Peace. The table reports average marginal effect estimates with standard errors clustered at the session level. All regressions include period fixed effects. * p<0.1; ** p<0.05; *** p<0.01

Next, we include *OneAlliance*, *FullConnect*, *maxAttack1* and *maxAttack2* in the same regression to investigate the relative importance of these state variables for each second. Table E5 reports the estimates, showing that while *maxAttack1* and *maxAttack2* tend to be more influential in earlier seconds, *OneAlliance* and *FullConnect* tend to take over the predictive power in later seconds. While some of these variables might overlap to an extent that prohibits a very precise interpretation, the overall result suggests that while intermediate state variables such as *maxAttack1* and *maxAttack2* are good predictors of Bully networks, it is eventually the stabilized pattern of alliance that absorbs their predictive power and determines the final outcome. We explore the dual dynamic process of attacking and alliance formation in more detail in the next subsection.

Table E5: Probit model estimates of network state variables (alliance and attacking variables)

	3 sec	4 sec	5 sec	6 sec	7 sec	8 sec	9 sec	10 sec
OneAlliance	-0.014	0.054	0.112***	0.109***	0.175***	0.201***	0.169***	0.214***
	(0.058)	(0.040)	(0.036)	(0.027)	(0.038)	(0.036)	(0.027)	(0.034)
FullConnect	0.215	-0.069	-0.188**	-0.282***	-0.332***	-0.341***	-0.397***	-0.372***
	(0.159)	(0.100)	(0.074)	(0.058)	(0.049)	(0.044)	(0.027)	(0.033)
maxAttack1	0.437***	0.375***	0.252***	0.171**	0.107	0.018	-0.059*	-0.039
	(0.060)	(0.062)	(0.070)	(0.069)	(0.065)	(0.056)	(0.034)	(0.046)
maxAttack2	0.498***	0.502***	0.447***	0.320***	0.208***	0.143***	0.059	-0.004
	(0.117)	(0.074)	(0.075)	(0.079)	(0.070)	(0.049)	(0.040)	(0.045)
N	725	725	725	725	725	725	725	725

Note: The dependent variable is 1 if the final status of the group is Bully, and 0 if Peace. The table reports average marginal effect estimates with standard errors clustered at the session level. All regressions include period fixed effects. * p<0.1; ** p<0.05; *** p<0.01

Finally, we also examine an additional explanatory variable, *TwoAlliance*, indicating whether there are exactly two three-member alliances. This is the case in which all but one pair of members are friends. We are agnostic about the predictive power of this variable but would like to

know whether it has the same predictive direction as *OneAlliance* or *FullConnect:* it is possible that any pattern of alliances falling short of being fully connected would eventually lead to Bully networks; but it is also possible that *TwoAlliance* serves as an intermediate step toward Peace networks. To investigate, we first ran Probit regressions on its own for each second. The estimates are reported in Table E6.

On its own, *TwoAlliance* does not seem to have much regular significant predictive power on the final outcome. However, when we estimate its coefficient together with those of *OneAlliance* and *FullConnect*, *TwoAlliance* significantly negatively predicts Bully networks starting from the 5th second or so, with regularity. The estimates are reported in Table E7. It is also interesting to observe that *TwoAlliance* tends to soak up part of the previous explanatory power of *OneAlliance*, previously a very significant predictor of Bully networks, although the estimate of *OneAlliance* never becomes negative. It is probably because there is no longer any variable that is a strong period by period predictor of Peace networks when *TwoAlliance* is included (that is, unobserved variables tend to predict Bully networks). It thus becomes easier to predict Peace networks than Bully networks. These results suggest that the groups with almost mutual friends are likely to find a way to eventually keep the peace, whereas the ones with one and only one three-member alliance consistently lead up to a Bully situation.

Table E6: Probit model estimates of network state variables (two alliances)

	3 sec	4 sec	5 sec	6 sec	7 sec	8 sec	9 sec	10 sec
TwoAlliance	-0.017	-0.152**	-0.107***	-0.039	-0.081	-0.148**	-0.087	-0.120*
	(0.118)	(0.076)	(0.039)	(0.049)	(0.057)	(0.063)	(0.083)	(0.065)
N	725	725	725	725	725	725	725	725

Note: The dependent variable is 1 if the final status of the group is Bully, and 0 if Peace. The table reports average marginal effect estimates with standard errors clustered at the session level. All regressions include period fixed effects. * p<0.1; ** p<0.05; *** p<0.01

Table E7: Probit model estimates of network state variables (all three alliance variables)

	3 sec	4 sec	5 sec	6 sec	7 sec	8 sec	9 sec	10 sec
OneAlliance	-0.016	0.001	-0.002	0.006	0.028	0.055	0.083**	0.105***
	(0.072)	(0.056)	(0.042)	(0.027)	(0.027)	(0.042)	(0.036)	(0.038)
TwoAlliance	-0.019	-0.201**	-0.225***	-0.208***	-0.240***	-0.259***	-0.192***	-0.194***
	(0.131)	(0.099)	(0.043)	(0.038)	(0.048)	(0.044)	(0.049)	(0.038)
FullConnect	0.094	-0.366***	-0.493***	-0.535***	-0.551***	-0.514***	-0.482***	-0.452***
	(0.208)	(0.119)	(0.053)	(0.042)	(0.029)	(0.032)	(0.032)	(0.029)
N	725	725	725	725	725	725	725	725

Note: The dependent variable is 1 if the final status of the group is Bully, and 0 if Peace. The table reports average marginal effect estimates with standard errors clustered at the session level. All regressions include period fixed effects. * p<0.1; ** p<0.05; *** p<0.01

Appendix F. Dynamics of network formation in the between-subject experiment

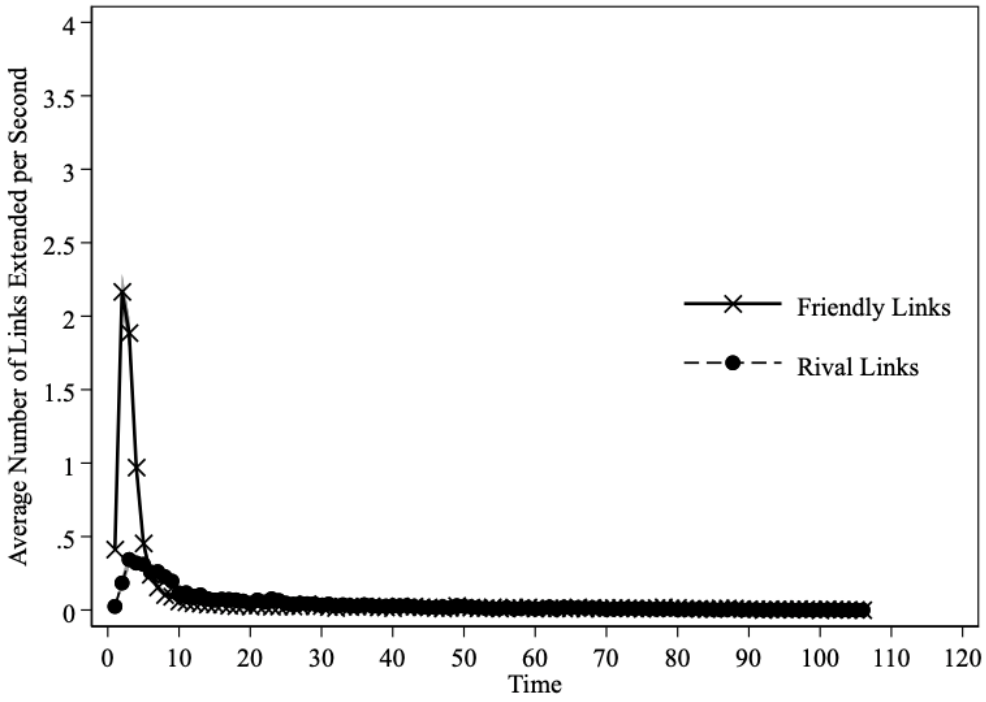


Figure F1: Extension of links per group, by second – between-subject experiment

Note: The grey shaded area indicates 95% confidence intervals.

Table F1: Percentages of 3-member alliances and fully connected networks, first 10 seconds – between-subject experiment

Time (seconds)	Peace		Bully	
	3-member alliance	Fully connected	3-member alliance	Fully connected
1	0.5	0	0.9	0
2	16.0	2.2	25.2	0.9
3	23.4	16.7	31.6	6.9
4	19.4	33.0	36.0	10.9
5	16.7	41.1	46.4	10.2
6	13.2	49.0	53.3	9.7
7	13.6	55.0	59.6	9.5
8	10.8	58.1	62.6	9.9
9	10.3	61.7	68.8	8.8
10	11.8	65.1	71.6	7.6

Note: "3-member alliance" is the situation in which three out of four players are mutual friends and the other player is a lone player; this is a necessary condition for Bully group. "Fully connected" is the situation in which all four players in the group are mutual friends.

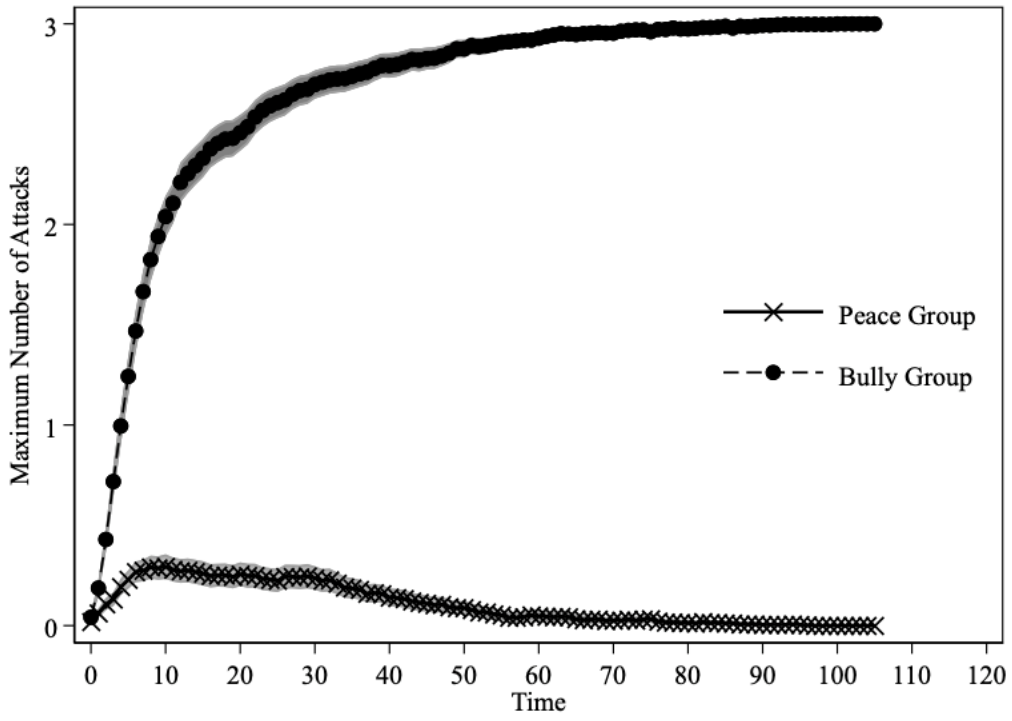


Figure F2: Evolution of maximum number of attacks received by any player per group – between-subject experiment

Note: This Figure shows the maximum number of attacks received by any player in a group. By definition, no player receives attacks at the end in Peace groups, and a player (final victim) receives 3 attacks at the end in Bully groups. The grey shaded area indicates 95% confidence intervals.

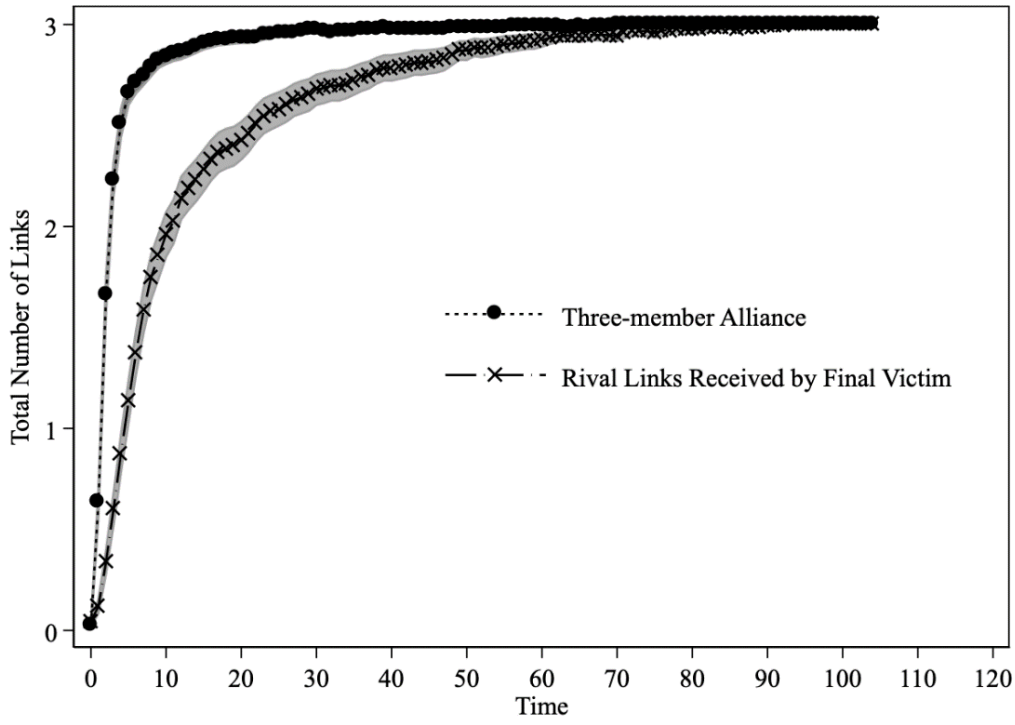


Figure F3: Evolution of average attacks received by the final victim and average effective friendships among the other three players, Bully groups – between-subject experiment

Note: This Figure shows dynamics in Bully groups. It plots the average number of enemy links received by the final victim and average number of effective friendships formed by the other three players (except for the final victim). The grey shaded area indicates 95% confidence intervals.

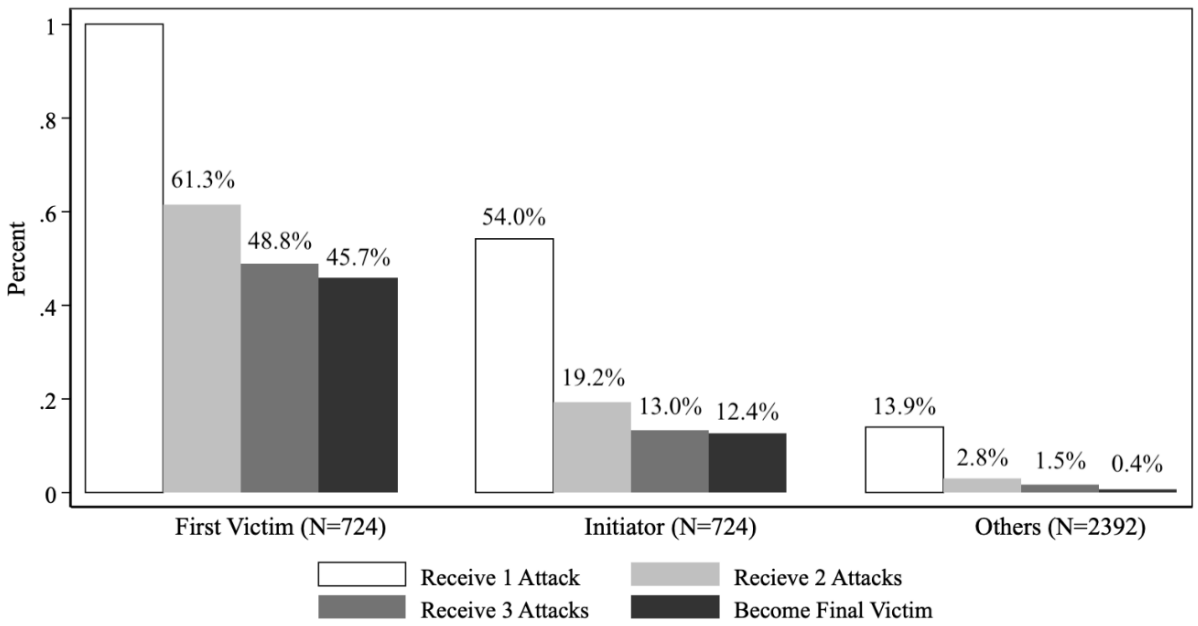


Figure F4: Transition to final victimhood – between-subject experiment

Notes: This Figure includes all 960 groups with a total of 3840 players. In 724 groups, there is ever a first victim. Correspondingly, there are 724 initiators of these first victims. "% Receive 2 (3) attacks" means whether the percentage of players who ever receive 2 (3) attacks during the whole round. "% final victim" means the percentage of players who become final victims.

Table F2: Random effects Probit model: determinants of final victims – between-subject experiment

	All groups				Bully groups			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
First victim	0.278***	0.279***	0.253***	0.278***	0.465***	0.490***	0.402***	0.473***
	(0.024)	(0.025)	(0.025)	(0.025)	(0.007)	(0.005)	(0.013)	(0.008)
Initiator	0.139***	0.139***	0.139***	0.139***	0.188***	0.185***	0.184***	0.187***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.019)	(0.019)	(0.017)	(0.018)
First victim (attack back)		-0.003				-0.091***		
		(0.011)				(0.022)		
First victim (befriending activity, above median)			0.049***				0.134***	
			(0.009)				(0.027)	
First victim (more friends than initiator)				-0.000				-0.083**
				(0.022)				(0.033)
N	3840	3840	3840	3840	1732	1732	1732	1732

Note: The dependent variable is whether a player is a final victim (=1) or not (=0). The table reports average marginal effect estimates with standard errors clustered at the session level. * $p<0.1$; ** $p<0.05$; *** $p<0.01$

Table F3: Random effects Probit model: determinants of initiators – between-subject experiment

	All groups			Bully groups		
	(1)	(3)	(4)	(5)	(7)	(8)
L1.Initiator	0.086***		0.085***	0.125***		0.122***
	(0.020)		(0.020)	(0.024)		(0.024)
L1.First victim	0.030		0.030	0.050**		0.049**
	(0.020)		(0.020)	(0.023)		(0.023)
L1.Final victim	0.041***		0.041***	0.038		0.038
	(0.014)		(0.014)	(0.031)		(0.031)
BNT score		-0.001	-0.004		-0.006	-0.007
		(0.008)	(0.008)		(0.005)	(0.007)
SVO angle		0.013	0.007		0.039**	0.032
		(0.021)	(0.020)		(0.017)	(0.024)
Risk-taking		-0.004*	-0.004*		-0.006**	-0.007*
		(0.002)	(0.002)		(0.003)	(0.004)
Competitive		-0.004	-0.002		-0.007	-0.005
		(0.004)	(0.004)		(0.005)	(0.006)
N	3648	3840	3648	1700	1732	1700

Note: The dependent variable is whether a player is an initiator (=1) or not (=0). L1.First Victim, L1.Final Victim and L1.Initiator denote being a first victim, a final victim, and an initiator in the previous round, respectively. BNT score takes a value from 0 to 4 with a higher number indicating a higher level of numerical sophistication. SVO angle takes a value from 0 to 90 with a higher degree indicating a higher level of prosociality. "Risk-taking" is self-reported general attitude toward risk-taking in daily life on the scale from 1 (not risk-taking at all) to 7 (extremely risk-taking). "Competitive" is self-reported general attitude toward competitive in daily life on the scale from 1 (not competitive at all) to 7 (extremely competitive). The table reports average marginal effect estimates with standard errors clustered at the session level. * $p<0.1$; ** $p<0.05$; *** $p<0.01$

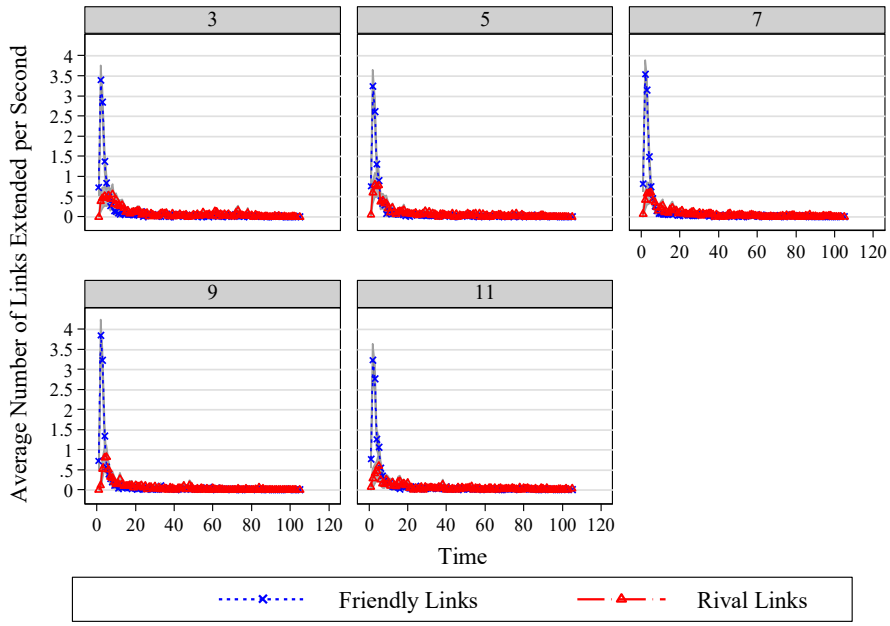


Figure F5: Extension of links per group per second by cost level – between-subject experiment

Note: The grey shaded area indicates 95% confidence intervals.

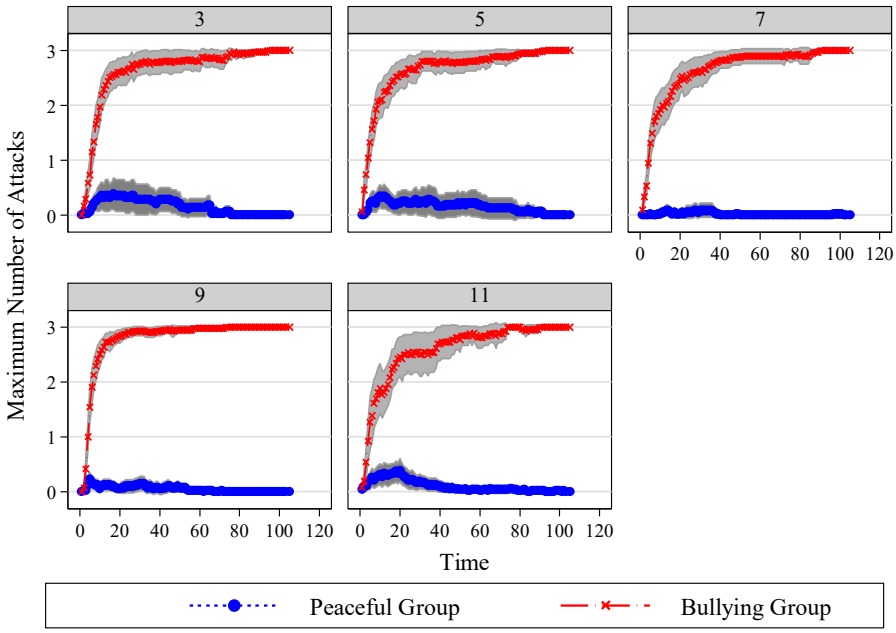


Figure F6: The evolution of the maximum number of attacks received by any player per group by cost level – between-subject experiment

Note: The grey shaded area indicates 95% confidence intervals.

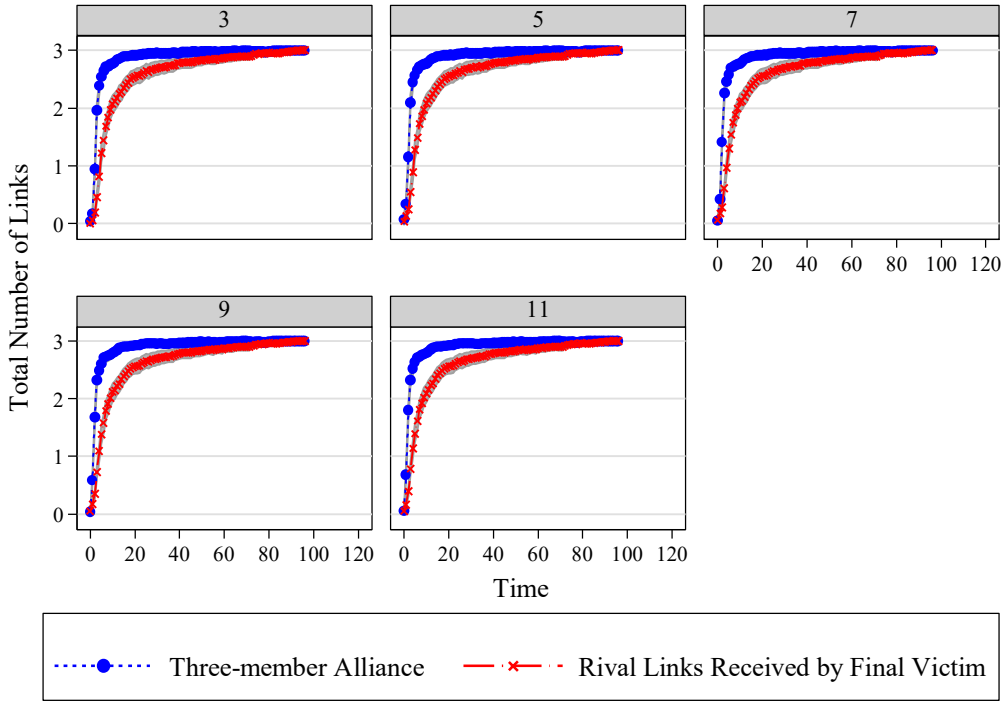


Figure F7: The evolution of average attacks received by the final victim and average effective friendships among the other three players in Bully groups by cost level – between-subject experiment

Note: The grey shaded area indicates 95% confidence intervals.

Table F4: The pattern of transition to final victims by cost level – between-subject experiment

Type	N	% Receive 1 attack	% Receive 2 attacks	% Receive 3 attacks	% Final victim
Cost=3					
First victim	286	100%	65.0%	57.0%	55.2%
Initiator	286	62.2%	24.1%	18.9%	18.2%
Others	708	21.3%	4.7%	2.5%	1.0%
Cost=7					
First victim	236	100%	61.9%	52.1%	49.2%
Initiator	236	48.3%	16.9%	13.1%	12.3%
Others	808	11.3%	2.0%	1.2%	0.1%
Cost=11					
First victim	202	100%	55.4%	33.2%	28.2%
Initiator	202	49.0%	14.9%	4.5%	4.5%
Others	876	10.4%	1.9%	0.8%	0.1%

Appendix G. A Quasi-dynamic analysis of network dynamics

In this appendix, we attempt to investigate the non-equilibrium dynamic interactions among players, with a particular focus on players' incentive to initiate an attack in an initial peaceful state. Given that we do not explore all possible dynamic patterns nor cover all possible initial states, we consider our analysis quasi-dynamic. The main purpose of this simple model is not to capture every player's strategy in real time accurately, but rather to highlight the coordination nature of the dynamics and hopefully inspire future theoretical research.

Given an initially peaceful state with 4 players, there are three possible final consequences, each of which can be supported as a reasonable equilibrium once an attack is initiated from player i to player j, with the other two players being bystanders n_1 and n_2. In Case 1, bystanders n_1 and n_2 successfully coordinate on supporting player i, resulting in a bullying outcome with player j as the final victim. In Case 2, bystanders n_1 and n_2 successfully coordinate on supporting player j, resulting in a bullying outcome with player i as the final victim. In Case 3, bystanders n_1 and n_2 cannot successfully coordinate and the attempt to initiate an attack fails, resulting in a peaceful outcome.

From the bystanders' perspective, since a successful coordination will lead to a payoff of $2k - c$ while a failed coordination provides a payoff of 0, players n_1 and n_2 have strong incentive to coordinate. However, given that decisions are made in a continuous-time setting, bystanders may not be able to achieve coordination in a timely manner. For simplicity, we assume that the two-bystanders make their decisions independently and simultaneously.[5]

Note that from a payoff perspective, the bystanders are indifferent between Case 1 and Case 2, both of which deliver a payoff of $2k - c$ to players n_1 and n_2. Suppose that the initiator i believes that each bystander attacks player j with probability μ (Case 1) and attacks player i with probability $1 - \mu$ (Case 2). For simplicity, not attacking is considered a dominated strategy and thus occurs with probability zero. Thus, the initiator i's expected payoff will be

$$u_i = \mu^2(2k - c) + (1 - \mu)^2(-6k) + \mu(1 - \mu)(0) = \mu^2(2k - c) + (1 - \mu)^2(-6k),$$

where the terms $2k - c$, $-6k$, and 0 stand for the initiator's payoffs in Case 1, Case 2 and Case 3, respectively.

The *individual rationality condition* for initiator i requires that $u_i \geq 0$, that is, the expected payoff by initiating an attack should be at least no less than staying in the initial peaceful state. This condition is equivalent to the following inequality based on the payoff function parameters in our game:

$$\mu \geq \mu^*\left(\frac{c}{k}\right) \equiv \frac{6 - \sqrt{12 - 6\frac{c}{k}}}{4 + \frac{c}{k}}.$$

[5] We have also considered an alternative setting where the two-bystanders make decisions sequentially. The main results remain the same and are available upon request.

A player may choose not to be an initiator even if initiating an attack is profitable. This can happen when letting someone else initiate the attack can potentially bring a higher payoff. By not initiating an attack, one can become a bystander, with 2/3 chance enjoying a higher payoff than being the initiator, while with 1/3 chance one may suffer from becoming the victim. Thus, the expected payoff for not initiating an attack will be

$$u_i^{'} = \frac{2}{3}[\mu^2(2k-c) + (1-\mu)^2(2k-c)] + \frac{1}{3}[\mu^2(-6k) + (1-\mu)^2(2k-c)].$$

The *incentive compatibility condition* for initiator i requires that $u_i \geq u_i^{'}$, which is equivalent to the following inequality:

$$\mu \geq \mu^{**}(\tfrac{c}{k}) \equiv \frac{3\left(8-\frac{c}{k}\right)-\sqrt{3\left(8-\frac{c}{k}\right)\left(16-\frac{c}{k}\right)}}{8-2\frac{c}{k}}.$$

We can show that $\mu^{*'} > 0$, $\mu^{*''} > 0, \mu^{**'} < 0$ and $\mu^{**''} < 0$. Also note that $\mu^*(0) = \min \mu^* > \max \mu^{**} = \mu^{**}(0)$, which means the incentive compatibility condition is always satisfied as long as the individual rationality condition is satisfied. We more formally state these results in the following proposition:

Proposition*: Given an initial peaceful state, the initiator's threshold belief on the bystanders supporting the initiated attack, denoted by μ^*, is an increasing and convex function of $\frac{c}{k}$.*

Proposition implies that an increase in the cost of initiating an attack leads to an increase in the threshold belief level regarding bystanders following the initiator, which all else equal, makes the initiation of an attack less likely. In addition, such an effect strengthens as the cost of initiating an attack rises. When $c = 0$, $\mu^* = \frac{3-\sqrt{3}}{2} \approx 0.634$; when $c = k$, $\mu^* = \frac{6-\sqrt{6}}{5} \approx 0.710$; when $c = 2k$, $\mu^* = 1$, which implies that initiation of an attack is not an optimal action under any possible belief if $c = 2k$.

In summary, by linking initiators' decisions to their belief about bystanders' behavior, our quasi-dynamic model provides a simple account of why some players are motivated to initiate an attack. The convex relationship between the initiator's threshold belief and the attacking cost (Proposition) is qualitatively consistent with the final networks observed in Figure 4, in that the higher the attacking cost, (the increasingly higher the threshold belief, and) the increasingly lower likelihood of a bullying outcome. It also appears to be consistent with our data showing that both the frequency of initiations of an attack (i.e., the proportion of initiators) and the likelihood of first victims becoming final victims tend to decrease with the attacking cost, especially when the cost increases from 9 to 11 (see Table D1 in Online Appendix D).

The quasi-dynamic model also provides an explanation for why first victims are most likely to be the final victim: since the value of threshold μ^* is always higher than 0.5 regardless of the attacking cost, an attack indicates that the initiator holds the belief that each bystander will join in attacking the first victim with more than 50% chance.

Furthermore, although we do not explicitly model the first victim's strategy and how that might influence the initiator's belief, the analysis provides a rationale for why first victims often fight back against initiators (probably as a strategy to escape from victimhood as discussed in Section 5.2.3). If the first victim fights back, the positions of the initiator and the first victim will become more symmetric. Since there is no payoff difference for bystanders between following the initiator and supporting the first victim, a more symmetric position should make the strategy of following the initiator less salient.

Finally, since the model assumes that the initiator will be the target of the bystanders with some probability (which is consistent with our earlier observation in Section 5.2.2), the expected payoff of a bystander is always higher than the expected payoff of the initiator. This prediction is also borne out in the data: among all groups in which attacking ever happened, initiators received on average 69.7, while bystanders earned on average 76.6. The difference is statistically significant at the 5% level using the Wilcoxon signed-rank test with session average as the unit of observation.

Overall, we view our simple quasi-dynamic model as a first step toward a better understanding of the rich dynamics observed in our network formation game. It would be valuable to explore further how to model players' coordination behavior more fully. For example, the continuous-time setup naturally calls for a model allowing for endogenous timing decisions by bystanders, whereby bystanders decide both when and whom to attack. An even more complete version should also endogenize the timing decision by the initiator given the observation that the initiator expects to earn less than bystanders.

We also do not necessarily consider coordination failure as the sole reason for not reaching a bullying outcome as the current model implies. The frequently observed peaceful outcome could be a result of players' other-regarding preferences such as inequality aversion. These preferences can lead players to prefer a Peaceful network in which everyone earns the same over a Bullying network in which a substantial payoff gap arises between the alliance members and the final victim.