

RESEARCH

Open Access



# Trust development in online competitive game environments: a network analysis approach

Anna Lena Fehlhaber<sup>1</sup> and Usama EL-Awad<sup>2,3\*</sup>

\*Correspondence:  
usama.el-awad@uni-bielefeld.de

<sup>1</sup> Leibniz University Hannover,  
30167 Hannover, Germany

<sup>2</sup> Faculty of Psychology  
and Sports Sciences, University  
of Bielefeld, P.O. Box 10 01 31,  
33501 Bielefeld, Germany

<sup>3</sup> Faculty of Medicine,  
University Clinic of Psychiatry  
and Psychotherapy Luebbecke,  
Ruhr University Bochum,  
44801 Bochum, Germany

## Abstract

Trust seems to become established even in scenarios where the prerequisites for trust are complicated by conditions that evoke scepticism. Nonetheless, trust emerges, a phenomenon that is to be comprehended and examined in the present experimental inquiry. In order to comprehensively capture the process, a competitive online game environment was used to document the development of trust networks, directionality, and strength using network analysis. Despite the conditions conducive to distrust in this game setting, acts of trust were exhibited. Robust trust bonds persisting over the course of gameplay appear to manifest mostly dyadic or triadic, with participant embeddedness within the network and homophily in terms of general trustfulness towards strangers being conducive factors for trust bonding and game survivability. This study hence contributes to the overall understanding of online trust development and offers several further research opportunities in a mostly unexplored field.

**Keywords:** MMSNA, SNA, Trust measurement, Trust, Behavioural experiment, Online environment

## Introduction

Trust is a complex and multifaceted concept which represents one's confidence in another party's ability to fulfill promises or expectations—not only in interpersonal relationships, but also in broader contexts such as trusting in systems, environments, or institutions (Mayer, et al. 1995; Ferrin, et al. 2007; Mcknight and Chervany 2001). Trusting someone puts you in a vulnerable position and the decision to trust necessarily involves taking a risk (Alós-Ferrer and Farolfi 2019; Coleman 1990).

The study of trust in personal interactions has a longstanding tradition, aiming to understand the reasons and processes behind individuals' willingness to engage in trusting relationships (Rotter 1980; Rousseau, et al., 1998; Sztompka 1999; Sztompka 2006; Gambetta 1988). In this context, empirical evidence has demonstrated the significant influence of various factors on the trust process. Specifically, the perception of trustworthiness, perceived familiarity, the way of communication, as well as social norms, expectations, and past experiences play crucial roles. These factors not only shape the trust

process but also activate mechanisms such as reciprocity, which can help to maintain trust in the future.

However, given the rapidly growing importance of the digital sphere, a focus was also placed on the study of trust and trust-building processes in the online domain. Several studies have examined the factors that contribute to trust building in online environments, considering the absence of traditional social cues that might indicate trust, such as gestures and facial expressions (Lewicki and Wiethoff 2000; Mcknight and Chervany 2001; Resnick and Zeckhauser 2002; Wang and Emurian 2005; Lewicki, et al. 2006). Most notably, quality of interaction, safety of the environment, social identity and self-expression, perceived similarity, and reciprocity were described as critical factors for trust. Nevertheless, in these studies, trust processes are predominantly studied in conventional e-commerce environments, focusing only on factors which are correlatively related to trust building such as opportunities for individuals to engage in interactions, to rely on reputation mechanisms, or to trust platform guarantors in the event of fraudulent activity (Liu and Tang 2018; Abrahao, et al. 2017; Choi and Leon 2023; Hieronana and Nugraha 2021; Duradoni et al. 2018, 2021). However, in contexts characterized by ambiguity and a lack of legal enforcement mechanisms, there is still a great need for more detailed and process-oriented longitudinal studies: This especially applies to the trading of digital gaming resources bypassing the official infrastructure, and moreover for unmoderated file-sharing platforms where the operators explicitly distance themselves from the content and any responsibility for fraud, and unregulated marketplaces or social media platforms (Hart 2021; Moeller 2023). In many online scenarios that people experience in their daily lives, situations arise that require a gift of trust, but leave the person granting the trust in a high degree of uncertainty as to whether this gift of trust will be reciprocated (Corbitt, et al. 2003; Williams, et al. 2017). Particularly due to the (perceived) anonymity on the Internet and the non-binding nature of promises outside of structured communities or platforms, this risk associated with giving trust is omnipresent (Masclat and Pénard 2012). At the same time, many Internet phenomena show that this entrustment is given to strangers despite the high level of precariousness or fundamental mistrust, for example because the exchange itself is not legal or because there are neither contractual securities nor regulations that take effect in the case of fraud or abuse by others. Even in such ambiguous circumstances, scammers and social engineers appear to easily gain the trust of their victims, regardless of their lack of a good or any reputation (Watters 2009; Prashanth and Cleotilde 2018). So far, studies examining trust building processes under ambiguous configurations, typically rely on observations of dark web trading platforms (Przepiorka, et al. 2017; Andrei, et al., 2023; Norbutas 2020; Lacey and Salmon 2015). However, most of the existing literature is an investigation of existing states without shedding light on developmental features or internal mechanisms involved in the decision-making process of trust attribution under ambiguous conditions.

The dark web undergoes constant changes, with websites and forums appearing and disappearing as swiftly as they emerge. This not only makes conducting long-term studies difficult but also hinders the ability to predict where such spaces might emerge, making it challenging to capture their initial state. This is understandable given the unpredictability of real-world phenomena, but it represents a methodological limitation

when attempting to study the process from its inception. Furthermore, the anonymity and lack of transparency within the dark web present inherent difficulties in comprehending its structures and activities. Another challenge lies in capturing the motives and motivations of the subjects, who are often observed without explicit consent, posing not only ethical concerns but also making cooperation for research purposes questionable and, if it occurs, highly selective.

### **Examining trust building processes with digital trust games**

For the present study, an approach was chosen that allows for the unification of observation and experiment to improve the understanding of trust-building processes in a hostile online environment under ambiguous conditions. To achieve this objective, a well-established methodology for assessing trust was selected as the fundamental basis, specifically experimental trust games (Camerer 2003; Berg, et al. 1995; Alós-Ferrer and Farolfi 2019). Experimental trust games study trust and cooperation by observing how one participant decides to entrust resources to another participant, who then decides how much, if anything, to return (Dasgupta 1988; Berg, et al. 1995). In this framework of the trust game, a mixed-motive dilemma can be introduced that involves one or more individuals who are motivated to cooperate and compete with each other, as also occurs in many everyday situations (Schelling 1960; Dawes 1980; Bouncken, et al. 2015). If a cooperative strategy is chosen, this can consequently lead to trust (Ross and LaCroix 1996) or may have been caused by trust in the first instance (Dirks and Ferrin 2002). In these mixed-motive dilemma situations, it is particularly risky for actors to bestow trust, since the trust-bearing person has the possibility and also the motivation not to honor the trust that has been given (Walton and McKersie 1966; Komorita and Parks 1995). Hence, trust games can be utilized to simulate mixed-motive dilemmas. However, trust games have limited ability to represent the complex reality in which such trust situations arise on the Internet. While anonymity or pseudonymity between two players can still be captured well by such games, it becomes much more difficult to track multiple individuals and their interactions and their evolving social relationships with each other. To better represent these complex trust situations and more accurately model the reality in online spaces, we used a custom social deduction game with rules known to the players and that align with the fundamental assumptions of experimental trust games.

Different from conventional trust games, there is no Nash equilibrium (a situation in which no player can improve their payoff by unilaterally changing their strategy) in this scenario. In the context of this custom deduction game, players were afforded the freedom to decide on trust and defection – or neither – over the entire course of the game. Moreover, they could communicate unrestrictedly with each other without having to disclose their anonymity. The rules stipulate that the game can only be won when all other players have been eliminated. The players are aware of these rules and understand that others will potentially attempt to exploit their trust.

This paper aims to comprehensively capture the entire process of trust generation, commencing from an initial state characterized by relatively equal distribution of social capital, wherein actors possess the freedom to choose between assuming the role of

trust takers or trust givers. This approach endeavors to provide a fresh perspective on the phenomenon of trust among strangers in hostile online environments.

The following exploratory research questions need to be posed for this purpose:

1. What kind of game actions do players generally perform in this competitive online game scenario? Under the circumstances and rules generated by this setup, does trust emerge at all?
2. What are the (trust and mistrust) dynamics and trajectories of play revealed in the different runs? Are these dynamics replicable (same rules, different players)?
3. When does on-sided online trust emerge?
4. When does mutual online trust emerge?
5. Related to the allocation of trust, do players favor trust building with strangers of similar low or high self-claimed trust behavior (perceived similarity relationship)?

## **Method**

### **Summarizing overview**

Across all nine game runs including 101 participants in sum, 173 trust situations have emerged in over 4365.68 h of total playing time. Each subject was randomly assigned to exactly one run, which lasted for several hours, days, or even weeks. The duration of the game was largely influenced by the subjects' game decisions. Findings reveal that at least one, often even two or more participant(s) in almost all runs tried to trust (each other), which could be shown via the in-game actions and the verbal logs. In some cases, this trust was reciprocated, and bonds were formed and that despite the knowledge of the impending betrayal. The scores collected via questionnaires on their own trustfulness also showed an attraction of similar scores to each other.

In this way, the results could provide clues as to how trust takes place on the Internet despite conditions contrary to trust, why many social engineering techniques are successful despite potentially existing mistrust, and which (group) dynamic processes influence the decision whether and whom to trust.

### **Observe real-life behavior in a game**

The observation environment chosen is an adaptation of a social deduction game designed by Halfbrick Studios which is called Tank Turn Tactics never officially released. Rather than being a traditional scientific-experimental game, this is a social deduction online game design in which trust and mistrust dynamics can emerge with players performing various competitive, cooperative or self-referential actions in order to try to remain the last survivor on the playing field. For this purpose, they possess both life points (HP) and action points (AP) that they can utilize in order to accomplish this objective as well as the opportunity to communicate with their fellow players.

These types of social deduction online game can be suitable for mapping real-world dynamics, as has been investigated in previous studies on the use of language features as a function of role assignment and effects of the game environment on game behavior (Girlea, et al., 2016; Zhang, et al. 2022; Xiong, et al. 2017). Reproducibility is also given

by the clear rules and specifications of the game. Moreover, unlike purely observational studies, these games can be replicated more easily: clear rules and specifications allow for a comparison of individuals but also groups of the different game runs (Glaeser, et al., 2000).

Overall, observing behavior in a normal game can be scientifically justified as it provides a way to observe behaviors in a natural setting while maintaining control conditions to ensure reliability and comparability of results.

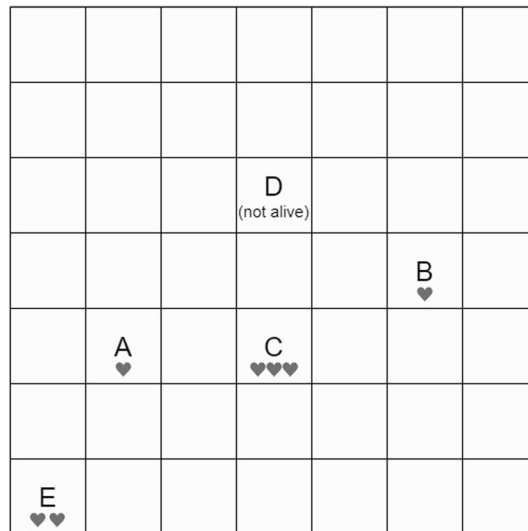
The reliability of the experiment was ensured by conducting several game trials under the same conditions, i.e., game rules, with different players. These conditions were slightly modified for three of the nine game runs in order to test whether and how contextual changes have an effect.

The validity of the measurement can only be estimated because while other, real game designs have already been subjected to the examination of scientific validity criteria, this unfortunately does not yet apply to Tank Turn Tactics. To increase internal validity, special care was taken to ensure that all participants received the same instructions which they also had to have understood in order to proceed, and that the overall conditions were as similar as possible. The external validity has to be considered as limited, since the self-selection of the participants (the invitation to participate was spread over various Reddit sub-forums) does not allow for a generalization. To test and increase construct validity, participants learned to take verbal logs prior to the experiment in a generic online board game and under the guidance of the experimental supervisor and used this method during the experiment to reveal their thoughts and reasoning. If no verbal protocols occurred for more than 30 s, subjects were reminded to continue verbalizing, also to ensure that as many considerations as possible were captured. These verbal protocols, in addition to formally defining trust as a transfer of resources without certainty of reciprocity, should allow for a test of whether the transfer of win-enabling game resources can be classified as trust decisions.

## **Data and measurement**

### ***Data***

A total of 129 participants were recruited via an announcement in multiple Reddit posts targeting millions of users, respecting online-study best practices (Kühne and Zindel 2020; Ho 2020). Of these, 101 participants took part in one of the game's nine runs conducted to which they were randomly assigned. Only rudimentary demographic characteristics were recorded so the participants could feel anonymous among themselves but also with respect to the study implementation which strove to minimize observational bias. The sample consisted of individuals aged 14 to 59 years from different places of residence all over the world, mainly USA and Europe. One common feature is the relatively high frequency of use of Reddit and Subreddit forums (one to four hours daily for most of them), which was intended as this study specifically tried to observe internet affine persons and their online trust behavior. The ability to maintain one's own anonymity was considered important in the study, in order not to influence the trust building process or the behavior of the participants. Therefore, no further sociodemographic variables were collected.



**Fig. 1** Example Board View

**Table 1** Cheat Sheet of Game Rules

Initial situation HP (health points): 3 AP (action points): 0 Range (action radius): 2	Effects per day (timing may vary): all players with 0 HP can vote to haunt a player · all players who are not haunted receive 1 AP · a heart is dropped at a random tile; a player moving on the tile gains 1 HP	Effects on defeat (reduction of a player's HP to 0): the AP of the defeated player goes to the one who conquered him/her · the defeated player joins the ghosts and can vote to haunt	Victory conditionally player but one have 0 HP
Possible Action	Cost	Target	
Move 1 square	1 AP	Self	
Shoot (reduce the HP of player in range by 1)	1 AP	another player	
Add a heart	3 AP	Self	
Upgrade range	3 AP	Self	
Revive	1 HP	Another defeated player	
Transfer 1 AP	1 AP	Another player	

*Notes: HP: Health Points, AP: Action Points*

Within the scope of this study, the games exhibited diverse durations, ranging from a few hours to several months, contingent on the evolving dynamics inherent to each game. Notably, these games operated in real-time, devoid of a turn-based structure, enabling participants to initiate actions at their discretion within daily, predetermined eight-hour windows.

Upon logging into the specially developed game, participants gained access to a comprehensive interface. This interface provided a holistic view of the game board, featuring the player's own position as well as those of fellow participants (see Fig. 1). Additionally, a menu on the left-hand side offered access to critical features, including a repository of the game's rules (see Appendix: *Game Rules*), individual and group chat functionalities, and a direct communication channel with the experiment coordinator.

Within the game environment, participants possessed the capacity to engage in various actions, contingent upon their available action points. These actions encompassed movements within the virtual landscape, strategic attacks against opponents, and the transfer of resources among participants, with a detailed breakdown available in Table 1. The players' objective was to maintain a playable character on the game board for as long as possible, ensuring that the character retained at least one hit point until the end.

Six seasons were performed according to the standard rules (see Table 1 for a very concise summary + Appendix: *Game Rules*), three more according to adapted quick game rules (see Table 1 + Appendix: *Game Rules*), in which time pressure played a crucial role which was modulated by the temporal frequency of allocation of action points.

### Measures

Trust attitude as well as trust networks were to be measured and modeled. *Trust attitude* was assessed using a scale for trust in strangers, which was developed and validated based on the Socio-Economic Panel and is called SOEP-trust. This scale is based on the conventional and well-known WVS-Trust, but expands upon it to improve validity and reliability (Naef and Schupp 2009 and Appendix: *SOEP-Trust*). *Trust networks* were formally assessed by fundamental network analysis techniques to observe emerging trust relations in in-game interactions, chat logs and verbal protocols.

The combined procedures for Mixed Methods Social Network Analysis (MMSNA) should ensure that not only the general structural characteristics of the network are captured, but also the quality of the relationships that emerge between the actors (Jack 2010; Froehlich, et al. 2020a, b). Therefore, not only the quantifiable interactions were recorded and categorized, but the communication between participants was also gathered and processed through a text-analytical summary. In addition, verbal thought protocols of the participants, which were rehearsed with each participant individually beforehand, were recorded during the whole game. When subjects logged into the game, the entire session, including their microphone recordings, was captured and subsequently transcribed. If necessary, the experimenter could remind the subjects in real time by private message to think aloud if the participants did not verbalize for several seconds while being in-game.

The analysis was conducted by stratifying the outcomes into the delineated categories as presented in Table 2. Solely the actions categorized as "intentional" were incorporated into the quantitative analysis.

**Table 2** Classification Categories for User Actions

Category	Definition
Accidental	Actions performed without conscious intention, often as a result of misoperations, misunderstandings of the user interface, or slips
Exploratory	Actions conducted with the intent of experimenting or exploring without a specific goal, aimed at understanding the functionality of a system
Intentional	Actions executed with a specific, goal-directed intention to achieve a concrete outcome or objective

## Models and analysis

### Model trust through network analysis

In this contribution, SNA is used for the description and analysis of the game and emerged game situations. Rooted in mathematical graph theory, social network analysis (SNA) can provide access to (social) relationship data. Interconnectedness between people allows for the inclusion of relationship context and relationship structure, which often cannot be included in the standard survey instruments (Knoke and Yang 2020). Social networks consist of the actors (nodes) as well as their relationships to each other (ties). With the ego network, as a part of the social network, the focus is placed on the individual, focal actor (ego) as well as its closeness to others (alter) (Şen, et al. 2016). Indeed, such social networks are also suitable for mapping qualities of the relationship between actors, as in the case of trust allocation. Regarding this experiment, the actors are the participants, each of whom can be modeled as nodes with relationships to other participants. From a network theory perspective and within the applied rules of the game, an exit network is formed, involving the withdrawal of different nodes (actors), leaving the remaining nodes available as potential trust recipients or trust (Hirschman 1970; Buskens 2002).

Using the game data obtained from the custom social deduction game, the social networks created between the actors were investigated. However, not only the game actions themselves are modeled as a network, but also trust decisions, which are named as such when a friendly game action was performed and this was not done accidentally but with awareness of the risk, validated using the verbal protocols of the participants.

In the past, personal trust within non-online contexts using social network analysis has been extensively studied for companies, within collaborations or households, as well as milieus, groups and between strangers (Ferrin, et al. 2003; Liu, et al. 2019). There is very little research utilizing social network analysis concerning the internet environment that does not focus on the organizational framework, such as commercial platforms (e.g. eBay etc.), or online social networks (e.g. Facebook, Twitter etc.) or alternatively, including not just some vague kind of relatedness between the actors, but manifested trust in the form of trusting actions. Moreover, even for non-online social network trust, it cannot be said with certainty whether the trust was entered into deliberately. This, in turn, is seen in scientific research as the crucial element to whether trust is established (Hargittai, et al. 2010; Resnick and Zeckhauser 2002; Coleman 1990).

Furthermore, it remains unclear to what extent the results of the non-online analyses about trust can be transferred to an online context. Thus, the presented research gap of self-organized systems on the internet in regards to the emergence of trust will be investigated in an exploratory way, including assumptions and findings for non-online contexts (such as effect of time and similarity between nodes) and using mixed method social network analysis (MMSNA) methodology with qualitative and quantitative network maps (Williams and Shepherd 2015; Froehlich, et al. 2020a, b) to model behaviors during the game and trust.

### Representation of complex social relationships using multidirectional graph models

In contrast to simple SNA graphs, complex graphs can acquire multiple relations with different directions (Koehly and Pattison 2005; Scott and Carrington 2011). These



multidirectional graph models provide a way to describe the observed system of trust actions in varying degrees of intensity. Additionally, multidirectional graphs can have different kinds of tiers, representing various attributes, such as interaction characteristics, to describe the nodes relationships (Pattison 1993; Shafie 2019).

To adequately represent game and trust decisions, two multidirectional networks are modeled. In a first network, all in-game interactions observable by other participants are systematically recorded and categorized. This classification encompasses actions that can be identified as hostile, neutral, or friendly, as delineated in Table 3. Actions that deplete the player's action points without being explicitly directed against other players are categorized as neutral. It should be noted that player movements were excluded from classification due to their frequent role in precluding either friendly or hostile interactions, thereby being indirectly accounted for within those categories. This approach was chosen in order to show all the relationships that were or were not established between the players.

In a second sub-network, only trust interactions that were performed during the game are modeled, to get a better perspective on the phenomenon being considered, which is trust under hostile conditions. Hence, only the friendly play actions validated as trust interactions through the verbal protocols were included in this second network. This design decision was made following an existing network analysis for trust interactions (Asim, et al. 2019), with trust behavior defined by (1) an action of transferring resources or reviving recorded through the game actions and (2) intention and perception, which was captured for both, trustee and trustor, with verbal protocols throughout the game.

### Visualization of game interactions and trust decisions

This entire network was calculated and displayed in its various iterations and states. After each trust interaction, the network and its structures were re-captured to map the evolution of friendly and hostile relationships between players at different points in the game. The number of interactions varied, in some cases significantly, depending on the time of day; on some days, no game decisions were made but only action points were collected and waited for, on other days the majority of actions were executed within several hours or sometimes, even minutes. Therefore, it was decided to divide the sequences according to the trust grants, opting against a time-based analysis approach.

In summary, the network is always a momentary snapshot, taken immediately after a trust action and has been placed in order to comprehend how the network emerges.

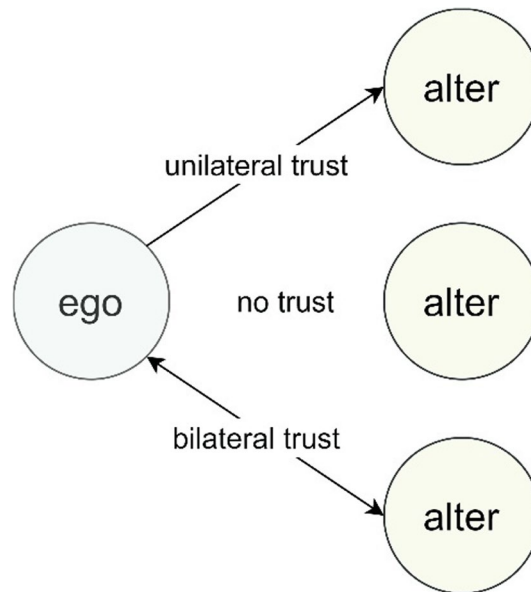
**Table 3** Classification of friendly, neutral and hostile actions

Hostile	Neutral	Friendly
Attack another player (only possible when alive)	Expand reach (self)	Give away action points (to someone)
Haunt another player to prevent him/her regenerating action points (only possible when dead)	Regenerate health points (self)	Give away health points (to someone)

*For more information about the full game rules, see Appendix: Game Rules*

**Table 4** Relationship matrix of trust network

	Ego does not trust	Ego trusts
Alter does not trust	No trust	Unilateral trust
Alter trusts	Unilateral trust	Mutual Trust



**Fig. 2** Focal node's relations

The Fruchterman-Reingold force-directed algorithm (also called anti-gravity approach, Fruchterman and Reingold 1991) is used to visualize each iteration of the network. It causes the distances between the nodes to change in the different iterations, since they repel each other if they are not held by a connection (i.e., if they are not linked together by an edge). Since in the present experimental design those graphs are drawn over the course of the different seasons, the evolution of the network and how it unfolds can be traced. As the state of the network is frozen after each cooperative iteration, there are as many visualizations as cooperative actions in the game run. A selection of salient network states is presented for each season, while the full evolution of the networks with all their states can be traced in the Appendix: *Degree Distributions and Main Networks*.

The other network to be investigated is that of trust decisions. It is a sub-network of the beforementioned overall network with its perspective placed on the individual actors involved in the network. The focal nodes ('ego') are directly connected to other nodes ('alter'), including directed ties to describe their relationship (see Table 4 and Fig. 2).

Only 'no trust' is excluded for visualization purposes from the sub-network since this is the initial situation for all participants: Since they don't know each other and because the rules indicate that trust can be very dangerous for one's own success in the game, general distrust can be assumed to be the starting point in this hostile-framed environment. However, unilateral and mutual trust as well as its relationship depth, represented by the number of associated interactions, are incorporated in this network.

### Statistical analysis

Numerous theoretical and empirical research contributions indicate that the approach of linking qualitative and quantitative methods leads to a more holistic perspective on social phenomena (Small 2011; Kolleck 2013; Froehlich, et al. 2020a, b).

Thus, quantitative as well as qualitative methods were used together. For capturing trust, formal quantitative SNA methods were utilized. To validate these assessments of in-game actions, Verbal Protocols as a variation of the qualitative SNA method of standard communication diaries as well as chat protocols were used to complement the former (quantitative) analysis and to validate whether it really was a trust interaction according to the theoretical definition, i.e. a conscious action. Furthermore, the SOEP-trust scale scores were modeled as node attributes for each participant to examine the homophily between them in the networks (Naef and Schupp 2009; Newman 2003; Foster, et al. 2010).

### Structural effects in network analyses

To examine the social network's structural effects, non-weighted node's in- and out-degree effects as well as network density and node's connection are considered. This will be done not only for the network of all interactions between the participants, but also for the subnetwork of trust, emerging in case of formula trust conditions.

This approach makes it possible to compare the two networks (complete interaction and trust-only network) with each other in terms of their connectivity. Connectivity and network density will be determined by their average degree, including the amount of strongly and weakly connected vertexes. A weak connection will be assumed when every vertex in the graph is reachable from every other vertex. To constitute strong connections, nodes must fulfill the aforementioned criteria and must be connected bilaterally.

These network's characteristics and arrangements are not only generated by the nodes' actions, but also may affect their further actions according to several theoretical and empirical studies (Gnyawali and Madhavan 2001; Jackson and Watts 2002).

### Effects of self-claimed trust behavior

Nodes can be examined in terms of similarity to determine whether, for example, similar nodes are more likely to form relationships with each other than dissimilar ones. Instead of using demographic information, which were allowed to be kept anonymous by the participants to simulate real world online situations, their trust towards strangers was modeled as node attributes. For this, an adapted and further improved scale for measuring trust in strangers (SOEP-trust, extending General Social Survey/World Value Survey trust) was used (Naef and Schupp 2009). With these items, the scale was modeled and used to characterize each actor's trustfulness as its node attribute. Accordingly, each node receives an attribute, which is the self-assessed confidence or trust in strangers, in order to examine later whether the (in)similarity of this self-assessment has an influence on whether edges form between the nodes.

## Results

The presentation of the results is structurally aligned with the research questions. After addressing Research Question 1 and offer a rough view of how players behave, a brief descriptive summary of each run ('season') is offered to address Research Question 2. Thereafter, Research Questions 3 through 5 will be analyzed in order to examine the contexts in which trust emerged during the game despite conditions that fostered mistrust.

### Research question 1: how the players operate during the game

To the author's best knowledge, this specific experimental game was applied for the first time in a scientific context. Accordingly, it must first be formally examined what actions players take in the first place and whether trust interactions are used—these are, after all, voluntary.

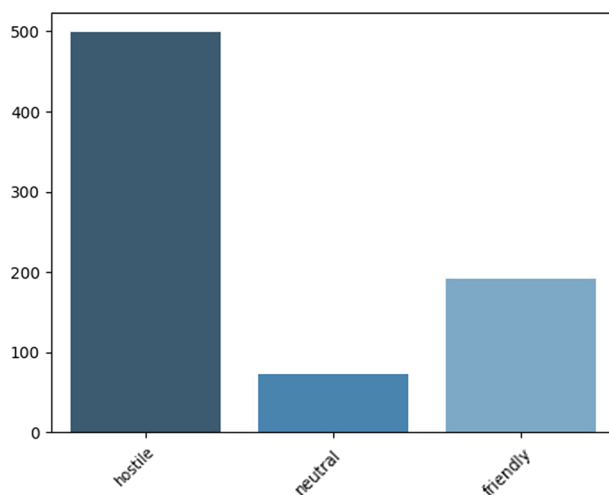
Referring to the classifications in Table 3, the interaction modes were organized and aggregated in Fig. 3. As can be seen, hostile interactions like attacks are executed with the highest prevalence. This is followed by friendly actions and finally action-point intensive neutral actions directed at oneself.

Even if the seasons are considered individually, the percentage distribution of hostile, neutral and friendly interactions is quite similar (see Fig. 4).

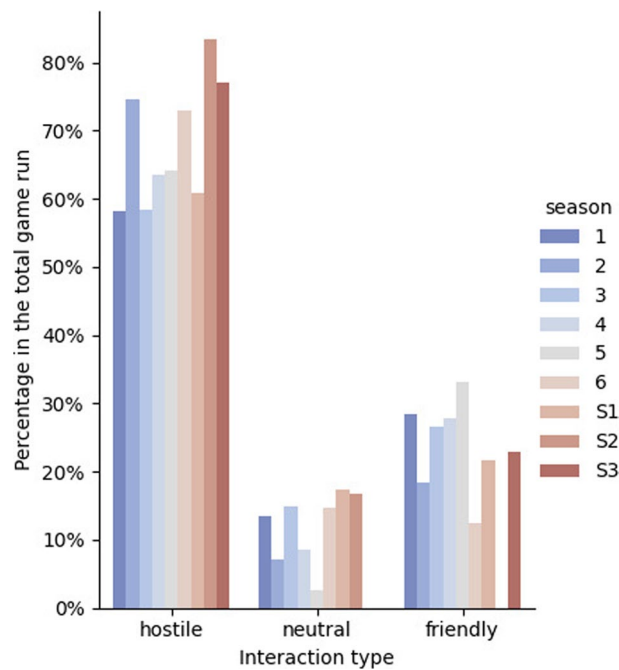
Trust, conceptualized as navigating uncertainty and embracing risk, was operationalized in this study through game actions involving the transfer of resources. To validate that it is a deliberate taking of a risk (and not an accidental action or the like), verbal protocols were used in addition to exclude accidental or unintentional resource transfers.

The friendly actions thus classified and specified to trust actions (in the case of the conscious decision for trust) were found in many of the game runs conducted, see Fig. 5.

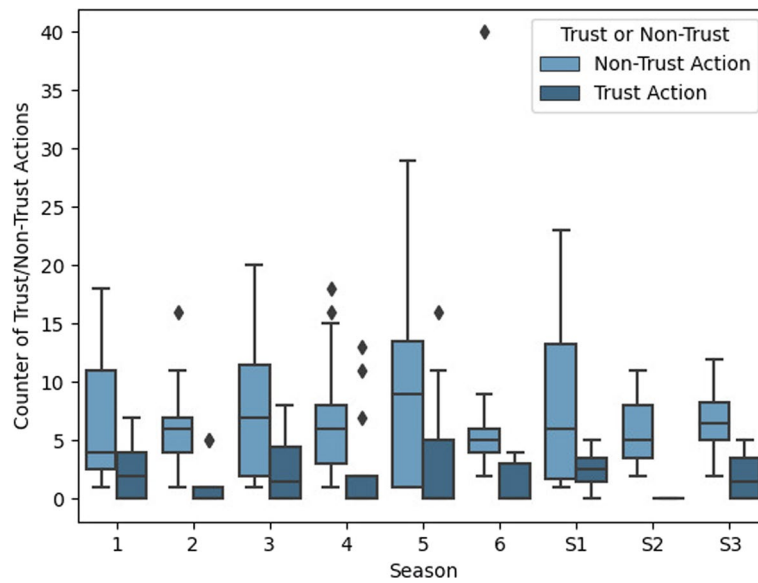
Thus, to summarize Research Question 1, overall hostile interactions dominate game play. The prevalence of neutral and friendly actions differs strongly between the game runs, also because these—unlike offensives—are not mandatory to win the game. With the exception of one game run, trust actions were evident.



**Fig. 3** Total number of categorized hostile, neutral and friendly in-game actions



**Fig. 4** Percentage of hostile, neutral and friendly interactions across all runs/seasons (standard seasons 1–6, quick seasons S1–S3)



**Fig. 5** Total Count of Trust and Non-Trust Actions across all seasons

Prior to answering the further research questions, a quantitative and qualitative sequential description of the nine playthroughs will be provided to get a deeper insight into the trust situations that occurred. After that, attention is turned to Research Questions 3–5.

### Research question 2: game dynamics and social interactions

Over the nine different runs, each has developed their own unique networks. The following describes the runs quantitatively and qualitatively. Not only the entire network, which includes the friendly, neutral, and hostile interactions, but also the network of trust, as well as the differences between the two networks and their connectedness – the higher the average degree, the more they are connected – are discussed as follows.

As explained before, the general network is presented for each run, respectively its different states during the game. This chronological sequence was visualized by numbering the different statuses (per run). In the network visualization, friendly, hostile and neutral/self-directed actions are mapped. Those friendly actions that could not be excluded as trust actions by consulting the verbal logs (i.e., it became apparent that it was an oversight or similar, but this happened only in very rare cases) are presented in the sub-network afterwards and show only the trust actions.

Players could choose a symbol for their character at the beginning to be distinguishable on the board. This symbol was assigned to a letter of the alphabet and serves as the designation for the nodes in the following. This letter is also used to describe the actions of the character in this run. It should be noted that the same letter can appear in different seasons, but behind it are different players, because each test person was only allowed to participate in one run.

#### *Season 1 – standard rules*

The overall network consists of eight nodes and 67 edges, with an average degree of 8.375 between the nodes. Its formation was visualized in Fig. 6. Overall, hostile actions clearly dominate, but friendly and neutral interactions can also be found between the players. An initial attack at the very beginning of the game is followed by a soothing, friendly gesture between player A and player C, intending to prevent further attacks according to A's verbal protocol. Also, B, who has previously attacked A, transfers resources to C and tries to convince C of an alliance. A then raises his/her stake for C and hands him/her more action points. Meanwhile, E and D also attack each other, D then starts a cooperation with H and shares the action points. H and D intensify their alliance, D eliminates enemies for H. In the end, the two allied players H and D peacefully agreed on who should win and decided to share the gain.

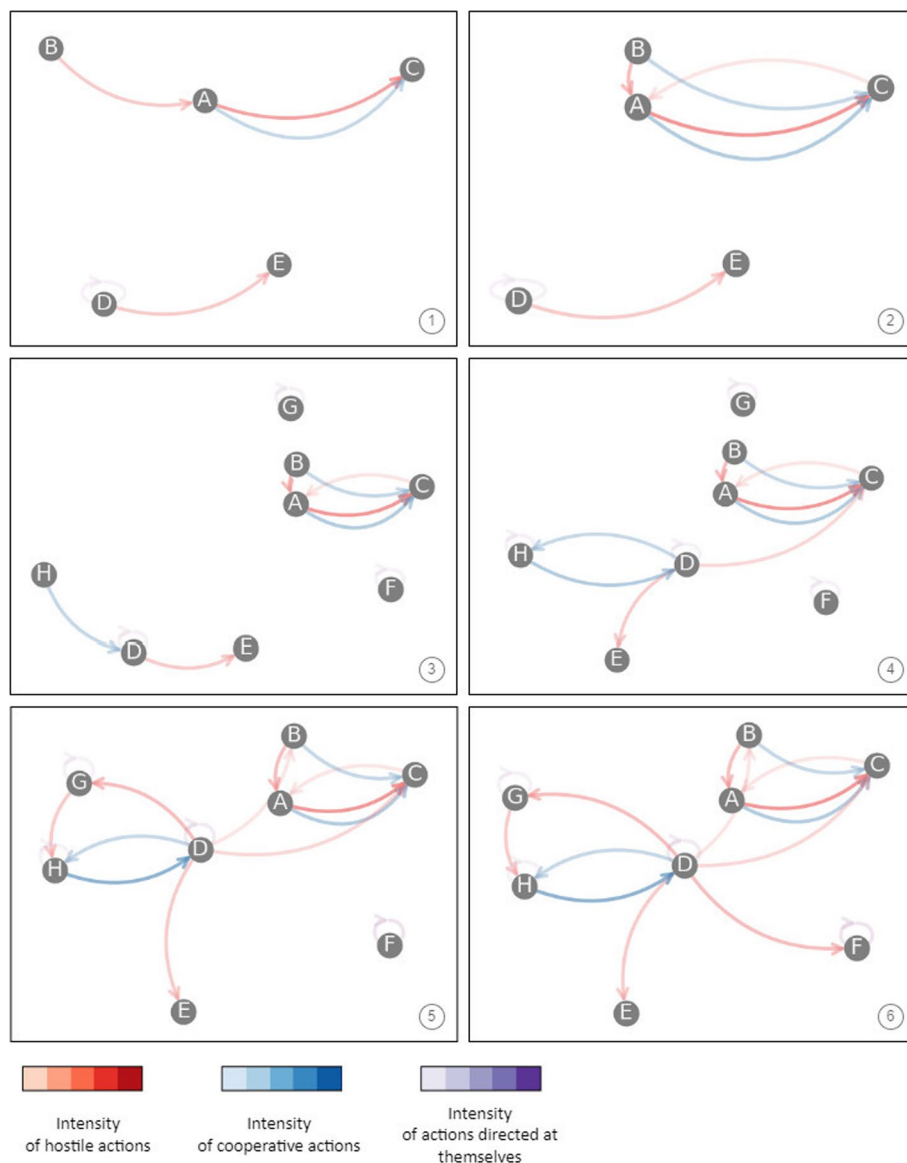
In sum, through these constellations throughout the game you will find five strong and two weak connected ties, these do not yet include differences between friendly neutral or hostile interactions.

Examining the trust-only network (Fig. 7), it becomes visible that five of the total eight nodes are involved in the sub network of trust. They have 17 edges with an average degree of 3.4, which is quite low compared to the connectedness of the overall network, explained by the fact that trust actions are quite costly.

#### *Season 2 – standard rules*

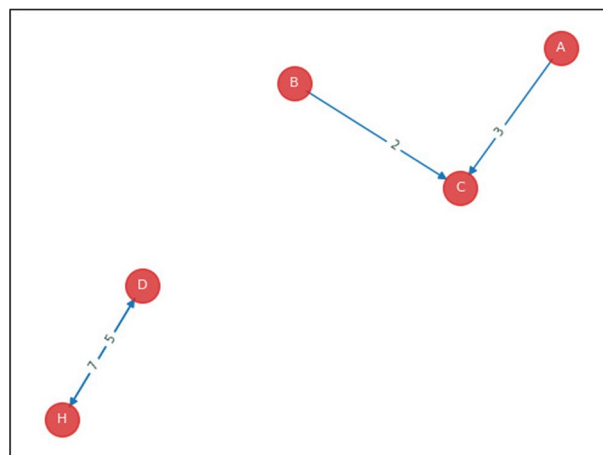
In the second run (Fig. 8), 12 other participants (nodes) established a total of 71 connections (edges). The average degree between these connections is 5.917.

At the very beginning, player assigned to letter I looked for allies by writing directly to four selected people. According to I, the choice was made in such a way that players



**Fig. 6** Salient states of the general network directly after a trust action, numbered in ascending order. Notes: Red: hostile, blue: friendly, purple: neutral/self-directed. Each node, including the letter, represents a game character

strategically well positioned on the playing field were preferred. After receiving a commitment from E to support him, he attacked two players directly one after the other until they were dead. E planned an ambush on I from the beginning and wanted to take advantage of the fact that I had taken such an offensive stance. To do this, E allied him-/herself with H. When I became aware of this, I offered to support E with resources—hoping for reciprocal behavior on the part of E. E accepted these and used them to then take I out. H and E then divided up the rest of the players on the board with the joint pool of HP and AP they gave each other as needed. When H betrayed and killed E at a time when H was supposed to be standing guard in case of attacks, the hitherto unapparent player G seized the moment to snatch victory by killing H.



**Fig. 7** Trust-only network for season 1. Notes: The number on the arrows indicates the number of trust interactions in each direction

In the sub-network of trust (Fig. 9), there are five nodes that have formed 12 edges with an average degree of 2.4. This is even lower than in the previous season. However, it is clear that the edge between E and H in particular, which is mutual, has a relatively high connectivity.

### Season 3 – standard rules

Over the course of season 3 (Fig. 10), there were 15 nodes which formed 120 edges. Many different strategies of cooperation and defection became apparent in this several months lasting playthrough.

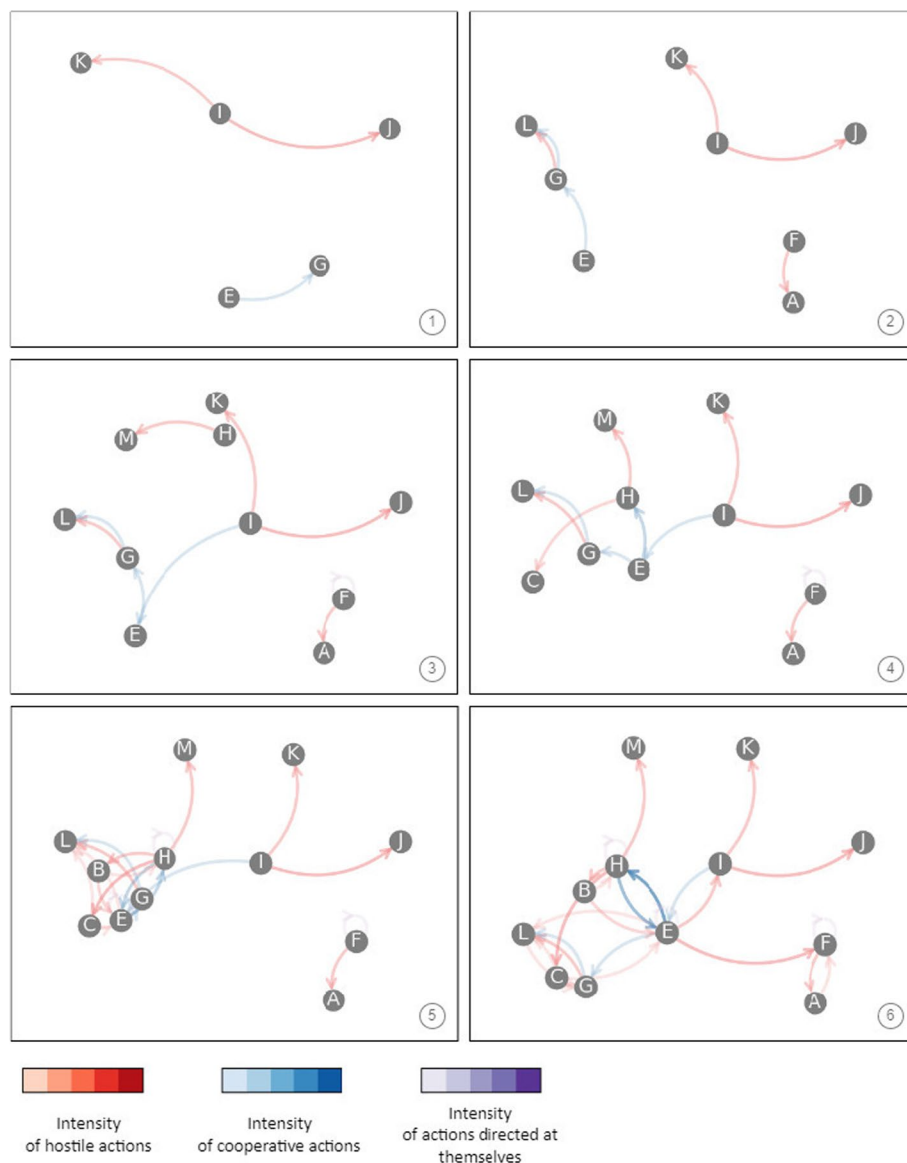
Player D and F made it clear at the beginning of the group chat that they would win the game and that whoever stood in their way would be out of luck. While D was massively supporting player F so that (s)he could commit attacks, D was additionally distributing resources to those teammates who were helping the two in their endeavor via player P. This resulted in a kind of resource distribution chain. Individual players like C, who tried to make alliances but were unsuccessful because many were afraid to publicly support through resource allocation if D and F could see it, not only remained alliance-less but also paid a kind of protection money.

More than half of the players (nine players) are involved in the trust network (Fig. 11), resulting in seven strong and two weak connections within the trust network. In contrast to the rather close connection of the main network (average degree: 8.0), the trust network shows a rather loose cohesion (average degree: 3.5), but a characteristic structure of passing resources along a chain.

### Season 4 – standard rules

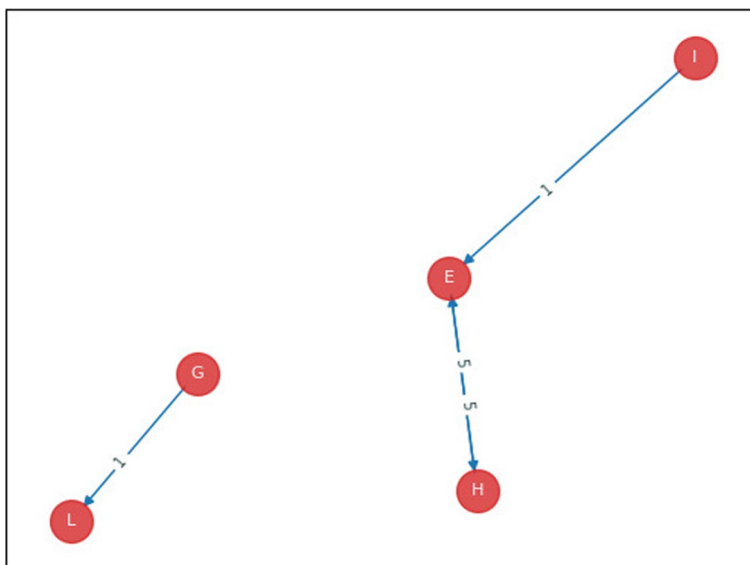
In this game (Fig. 12), 17 players participated, who in turn generated 129 edges. The average degree of these was 7.589 and included five strongly and one weakly connected node.





**Fig. 8** Salient states of the general network directly after a trust action, numbered in ascending order. Notes: Red: hostile, blue: friendly, purple: neutral/self-directed. Each node, including the letter, represents a game character

In this run, the network of alliances that had been built up broke down in between, but was—in part at least—re-established. Player P actively decided against any form of alliance, but terrified the other characters relatively early in the game. While some time initially passed with all players waiting or making deals among themselves, P used her action points to wipe out one player after another and get their action points transferred. After five successful kills by P, player U intervened after being assured of support by several other players. The agreement was that U would get one action point from each of the other players immediately after the attack against P, so that U could kill the player P. The other players would give U one action point. However, none of the pledgers wanted to keep the agreement due to misinformation player D spread, so that after several attacks on P, without being able to finally kill her, U took



**Fig. 9** Trust-only network for season 2. Notes: The number on the arrows indicates the number of trust interactions in each direction

flight and, in the future, no longer entered into cooperation with any of the other players. The weakened player U was then disposed of by one of the players who originally promised to support U.

Player D, who built up a close network of confidants right from the start and fomented distrust between the others via appropriate messages, claiming that the other party had already talked to him and wanted him to get rid of the player.

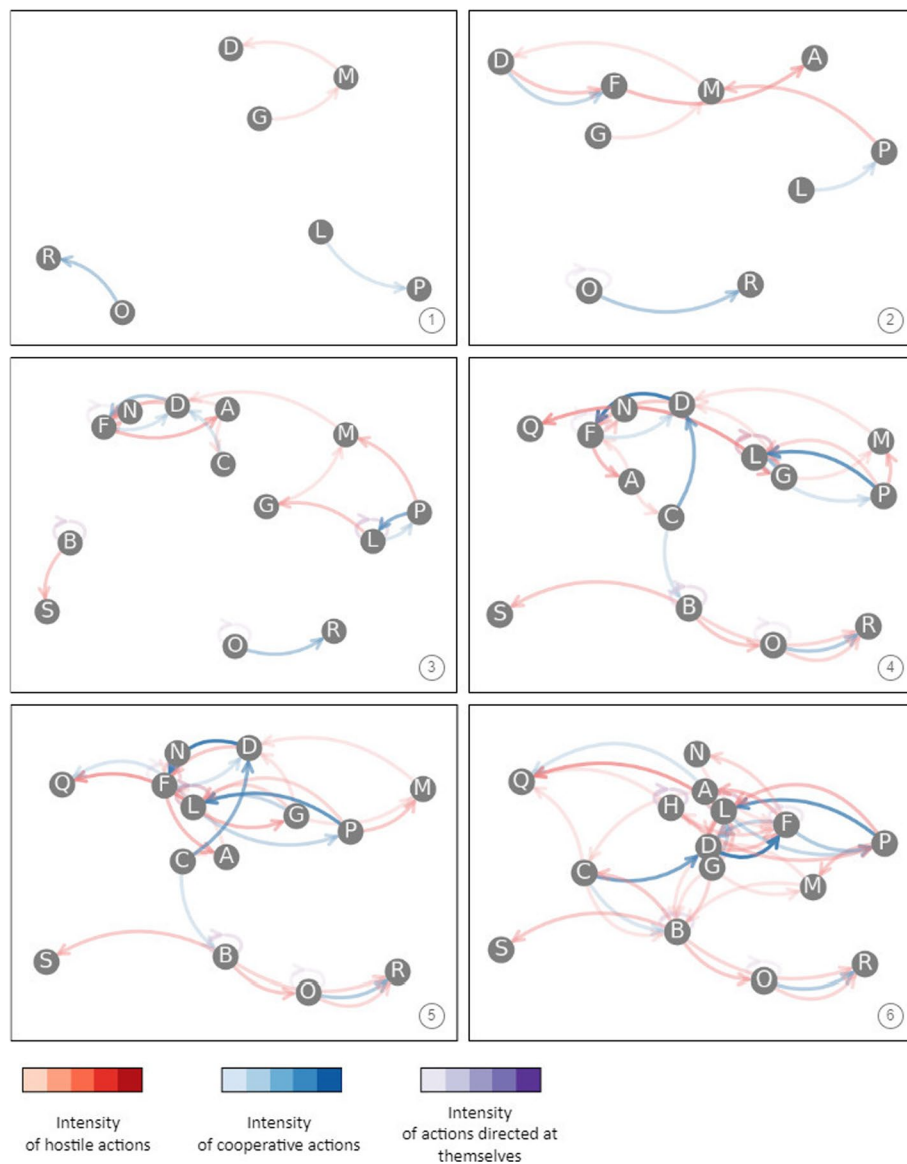
The subnetwork (Fig. 13) of trust consisted of six nodes, which is relatively low considering the number of players in total. Together they formed 35 edges with an average degree of 5.83. Player D, who took a central role not only in the overall network but also in the specific subnetwork of trust acts, ended up winning the game, probably also because of the prominent position he built up over several mutual relationships.

**Season 5 – standard rules**

The main network of this round (Fig. 14) consisted of 15 players (nodes) and 187 interactions (edges). With an average degree of 12.467, this is relatively high, which speaks for a strong interconnection of the nodes. These are relatively broadly distributed, so two of the components are strongly connected and one is weakly connected.

This season also did not see the formation of fixed alliances, as was the case in the game runs considered so far. Rather, individual alliances were formed that were quite fragile and broke down through betrayal – with exception of player Q’s concentric network.

Q was not only the player responsible for the most kills (seven kills, the next most by a single player in this run is just two), but also the player who was revived the most by allies. Using a mixture of bribery and threats, he secured the support of other players at the beginning. In particular, player Z, to whom Q sent an anonymous real money transfer in an amount larger than the game winnings, proved to be loyal and helped Q a lot.

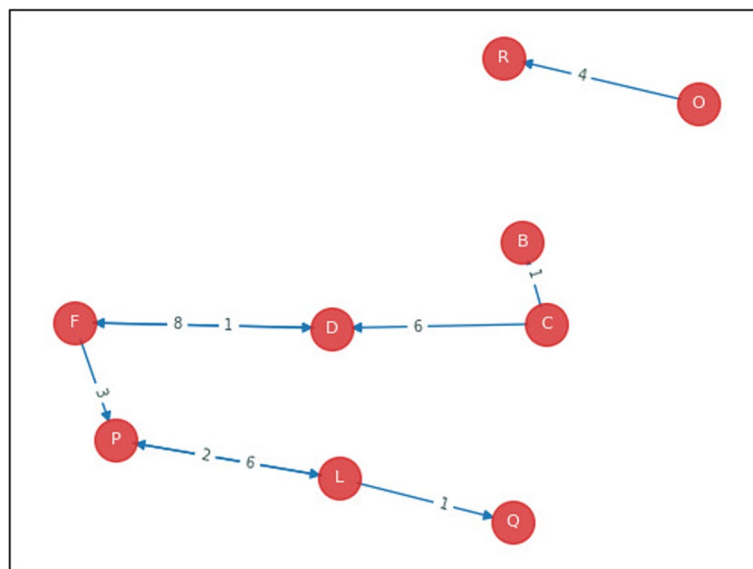


**Fig. 10** Salient states of the general network directly after a trust action, numbered in ascending order. Notes: Red: hostile, blue: friendly, purple: neutral/self-directed. Each node, including the letter, represents a game character

When only Q and Z were left at the end of the game, Z sacrificed him-/herself by giving away all his/her life points, including the last one, to Q so that he would win.

D and F tried to copy them, but their relationship of trust broke down when F refused to revive D.

A strong interconnectedness also emerges for the sub-network of only trust interactions (Fig. 15), in which just under half of the players (seven players) participated. They formed 47 edges, which led to an average degree of 6.714. Formally, several unilateral and mutual relationships are observable. It is particularly interesting that actors respectively their nodes sometimes maintain both, i.e. relationships to which they react with reciprocal actions and those in which they do not react or even knowingly betray trust.



**Fig. 11** Trust-only network for season 3. Notes: The number on the arrows indicates the number of trust interactions in each direction

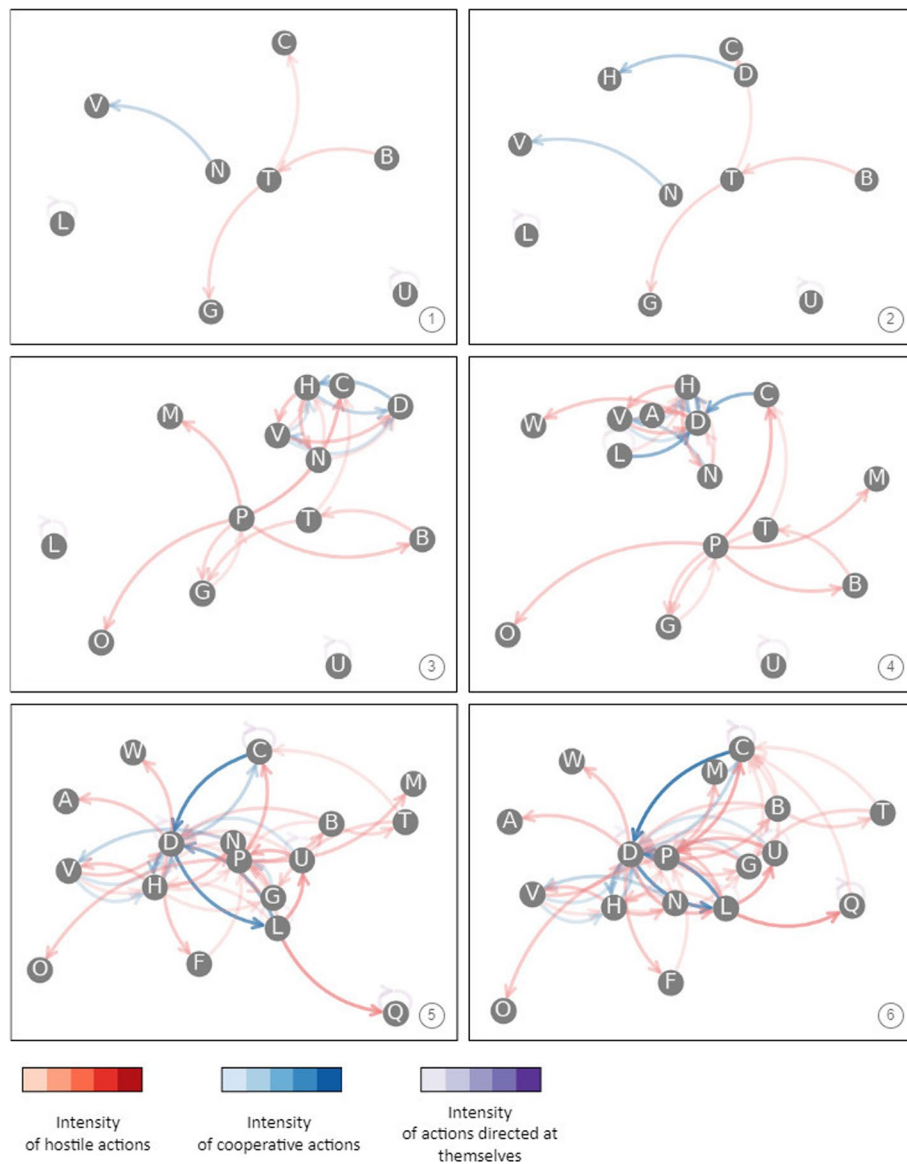
### Season 6 – standard rules

The total network (Fig. 16) consists of 18 nodes and 89 edges with an average degree of 4944. The number of strongly connected nodes is relatively high at 13, with one additional weakly connected node. The reachability from all vertices of the nodes is relatively high due to six strongly and two weakly connected components. Half of the players failed to generate any out-degrees, as they were killed relatively early by their co-players. This is also evident in the star-shaped arrangement of the network.

Players P, D, T, and F agreed on an alliance in which P would distribute resources and F would coordinate attacks. F also used action points for this several times for self-related actions, such as to increase their own range. Among themselves, P, D and T expressed concerns about this, but did not dare to confront F about it, because (s)he had so many action points and range in the meantime that (s)he could have been dangerous to them as well. E, who went on to win the game, ended up sharing the win with his/her alliance, thus honoring the trust placed in him/her.

In addition to this web of trust allocation, another relationship of resource transfer existed between C and X. C credibly explained to X that the latter could not win the game anyway, but that she urgently needed the money and would be glad of X's support to come to victory. However, this one-sided relationship was then ended by the superiority of the alliance of P, D, T and F. C ignored X, after X got useless for her, but tried to convince other players to help her.

The described gameplay led to the situation that one-third of the main network is part of the trust sub-network (Fig. 17): six players with 11 relationship ties, giving a relatively low average degree of 1.833. Mutual trust assignments could not be observed in this run.



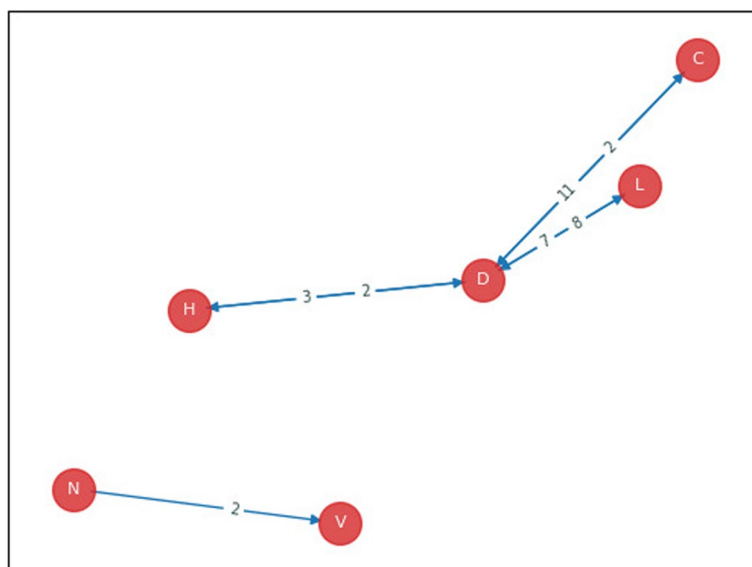
**Fig. 12** Salient states of the general network directly after a trust action, numbered in ascending order. Notes: Red: hostile, blue: friendly, purple: neutral/self-directed. Each node, including the letter, represents a game character

**Quick season S1 – quick game rules**

In this first game run under the quick play rules (action point drop: every minute instead of standard: daily, Fig. 18), seven nodes formed with 46 edges. The connectedness in the network, measured by the average degree was 6.571. Five of the nodes were strongly connected, one weakly.

Right at the beginning and before they could perform their own actions, 3 of the 7 characters were killed by the others.

Player D and player G were permanently online during the entire game and dominated the game. Shortly before the end of the game, G was even willing to transfer resources to D as an offer of peace between them after D revived another player (A).



**Fig. 13** Trust-only network for season 4. Notes: The number on the arrows indicates the number of trust interactions in each direction

D then killed the just revived player A again, saying that he wanted to win and insulting G and the—from his perspective – ‘pathetic’ player A who seriously believed that he could trust D. G then killed D’s ally (B) and not only won the proxy war, but also managed to keep D so busy in private chat sowing self-doubt, that D got careless for a few seconds and then was killed by G as well.

Four of the seven players had at least one trust interaction, and a total of 10 edges are present in the trust network (Fig. 19). The average degree was correspondingly 2.5. Three of the nodes were strongly bound, one weakly bound.

#### **Quick season S2 – quick game rules**

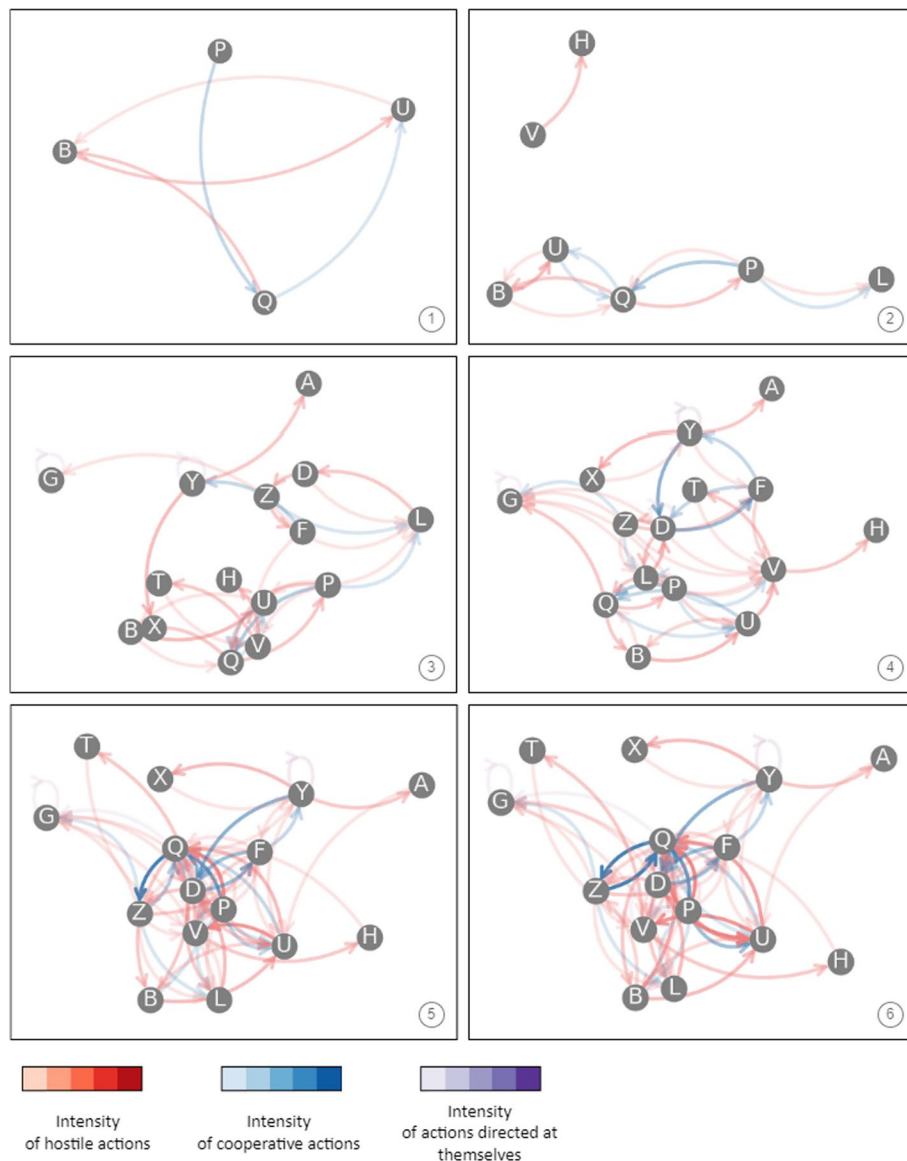
In this run, which also took place using the quick play rules (action point drop: every two minutes instead of standard: daily, Fig. 20), five players participated and formed 18 edges, giving an average degree of 3.6. After all, five components were strongly bound and one weakly bound. Trust interactions were not exercised during this run.

Also, no arrangements were made among themselves. Only after the game the winner expressed in the global chat that she was happy to have won. During the game, hostile actions were performed very frequently, but as mentioned, without any communication between the participants.

#### **Quick season S3 – quick game rules**

In the last quick game (action point drop: every hour instead of standard: daily, Fig. 21) four participants took part, forming 35 edges and thus having an average degree of 8.75. Only one strongly and one weakly bound component can be found in this relatively small network.

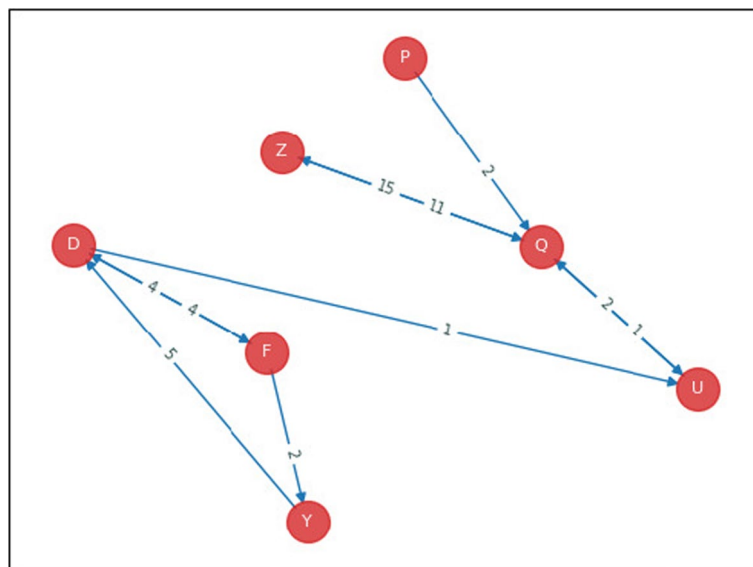
D started the game by attacking A directly. However, D did not have enough action points to kill A and panicked as a result. L intervened and got rid of player A after he



**Fig. 14** Salient states of the general network directly after a trust action, numbered in ascending order. Notes: Red: hostile, blue: friendly, purple: neutral/self-directed. Each node, including the letter, represents a game character

killed P. Player A then took the opportunity to haunt L, so that last-mentioned no longer regenerated any action points. L now had the problem of having no action points left and had to rely on the help of someone else if he still wanted to win the game. He then contacted P, who was suspicious. L promised to revive P and give her all the action and life points he had left, so that the others would at least not win. P agreed and L kept the agreement. P wanted to return the favor and, when he had enough life points, revive L and help her, but the other two players were too quick and killed them both.

Two of the four participants exercised trust interactions and generated as many as 8 edges, which is a relatively high intensity with an average degree of 4.0 (Fig. 22).



**Fig. 15** Trust-only network for season 5. Notes: The number on the arrows indicates the number of trust interactions in each direction

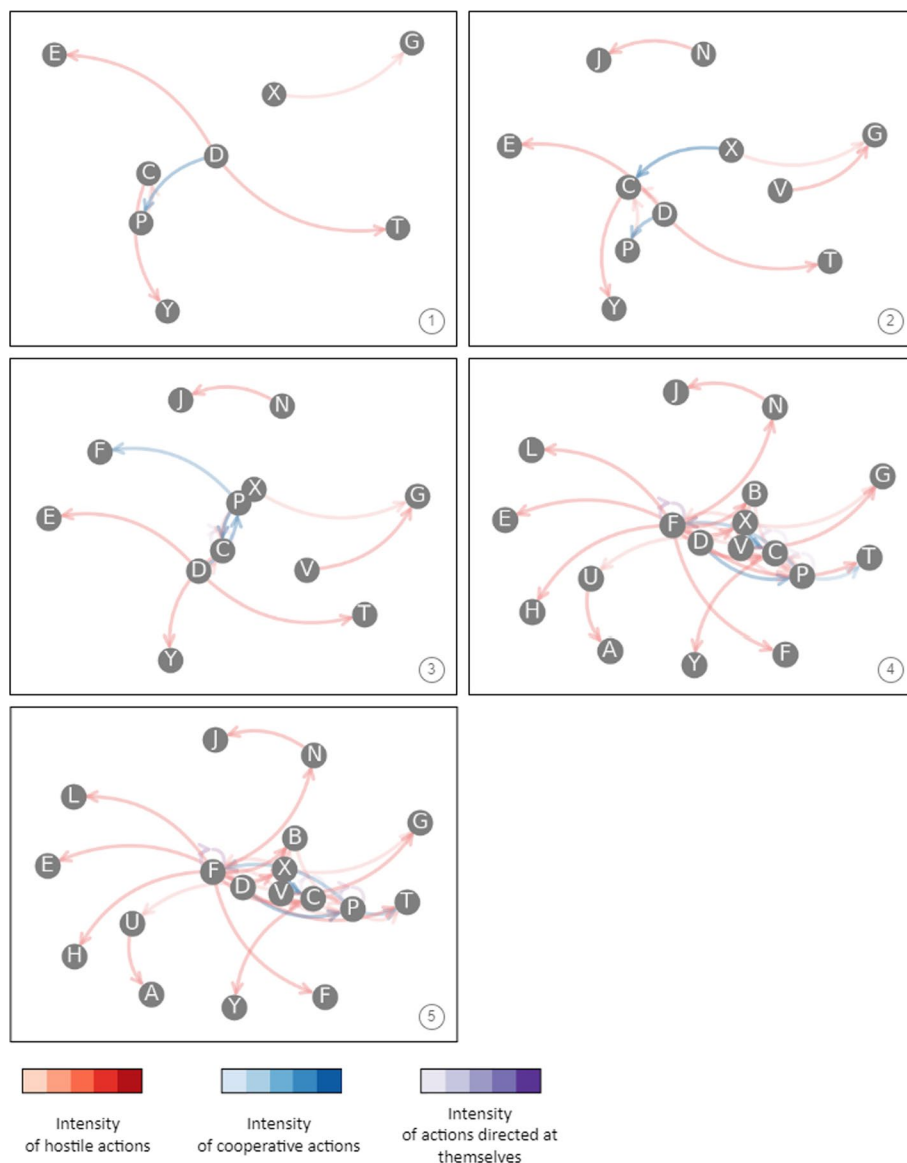
In answer to Research Question 2, it can be said that, among many other social phenomena, there were quite different dynamics of trust and mistrust to observe. In the nine different runs, it became apparent that trust emerged despite the adverse conditions, namely the rules of the game, which made betrayal indispensable. The actions of others often directly affected their own possible game actions, as well as opportunities for alliances or enmities.

#### Research question 3 and research question 4: when did unilateral and bilateral trust relationships emerge?

To answer Research Questions 3 and 4, the unilateral and bilateral (mutual) trust relationships in the different runs are examined in a comparative manner, theoretically founded on focal node's relation (Table 4 and Fig. 2). They are gathered and sorted in Table 5 and evaluated in proportion (low/medium/high). In the case of mutual trust, a distinction is made between (relatively) balanced and non-balanced mutual trust relationships, where identical take-and-give was considered balanced, a relative percentage difference of a maximum of 35% was considered relatively balanced and higher relative percentage differences were considered unbalanced. Between the actors between whom trust actions took place, the number of trust actions were summed for each trust action. For example, the first entry of the following table (Table 5) is to be interpreted as such: In season 1, two unilateral connections of trust were observed. In the first instance, two one sided gifts of trust were made, and in the second instance three unilateral trust actions were taken without any reciprocation. In addition, in the same season there was one mutual relationship relatively balanced with one player performing five trust actions upon an ally, that reciprocated with seven trust actions.

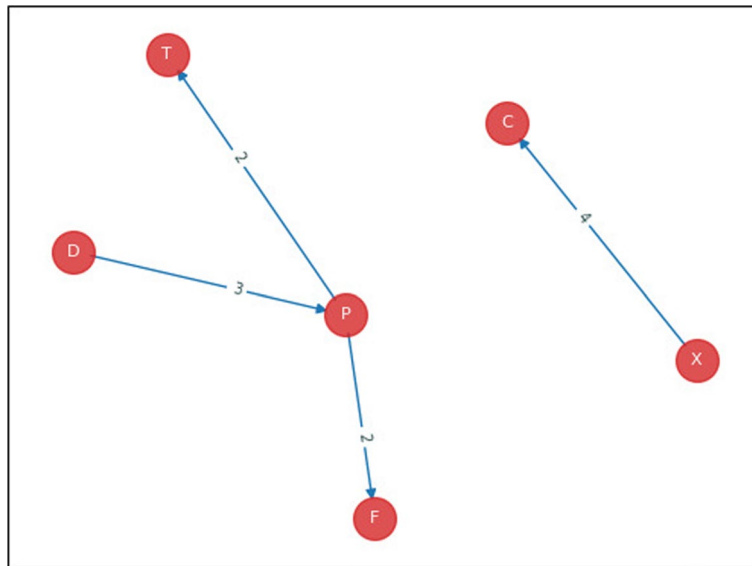
Both unilateral and mutual trust relationships emerged during play to jointly defend themselves, exchange resources, or signal peaceful intentions to avoid aggression. It is clear from the summary table (Table 5) that unilateral relationships have taken



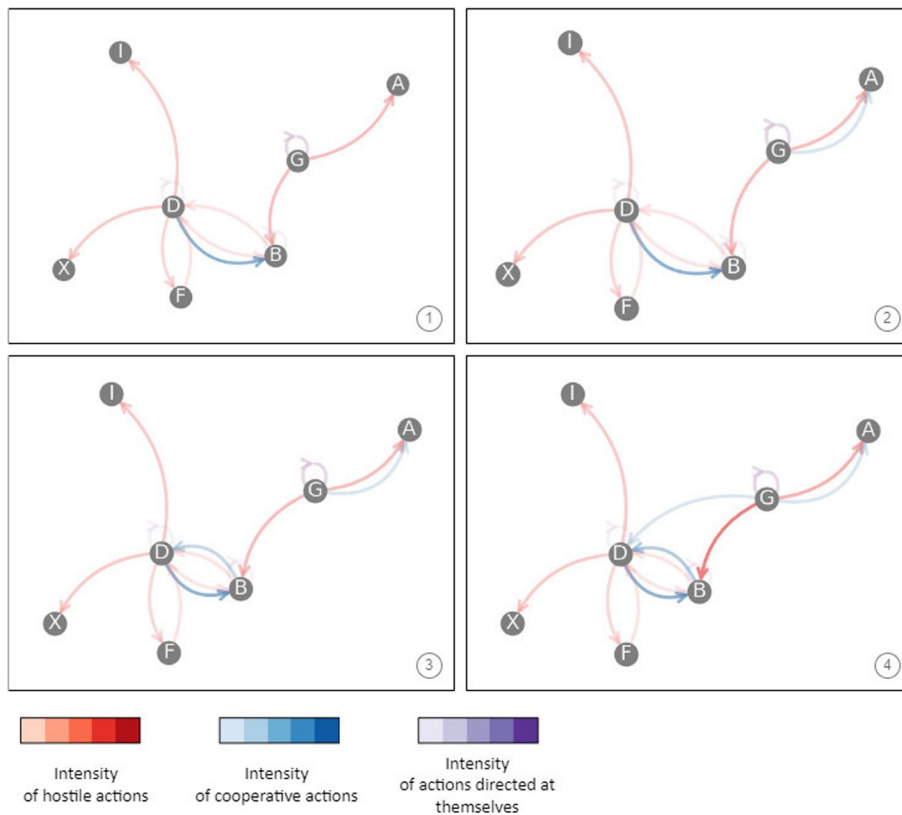


**Fig. 16** Salient states of the general network directly after a trust action, numbered in ascending order. Notes: Red: hostile, blue: friendly, purple: neutral/self-directed. Each node, including the letter, represents a game character

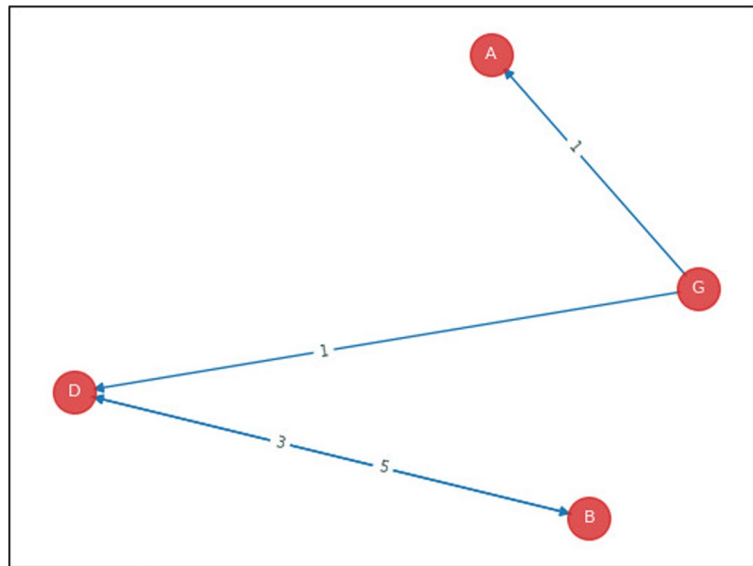
place in many of the runs. In most cases, these were characterized by low intensity, but in some exceptions, due to threats, bribes and (plausible) promises, they also reached high intensity without corresponding reciprocity. These phenomena could be explained by loss aversion and the high emotional investment (McAllister 1995; Lewicki and Wiethoff 2000), as well as risk-taking willingness (Serva, et al. 2005; Jøsang and Presti 2004; Boon and Holmes 1991) that occurred in these situations, as can also be seen from the verbal protocols: ‘Well, I’ve decided to trust [...], so I’m going to continue doing that now, and I did it before, too, because it would have been for nothing’ or ‘If he lies to me I’ve lost anyway’.



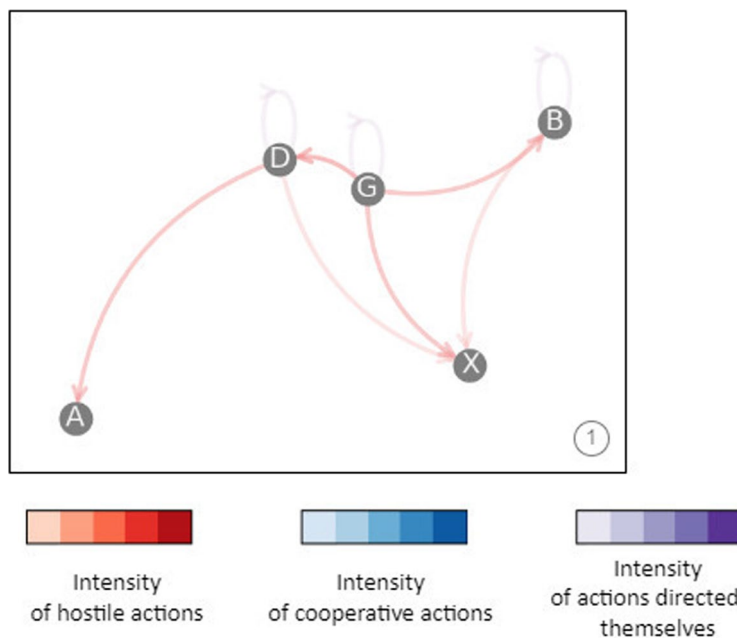
**Fig. 17** Trust-only network for season 6. Notes: The number on the arrows indicates the number of trust interactions in each direction



**Fig. 18** Salient states of the general network directly after a trust action, numbered in ascending order. Notes: Red: hostile, blue: friendly, purple: neutral/self-directed. Each node, including the letter, represents a game character

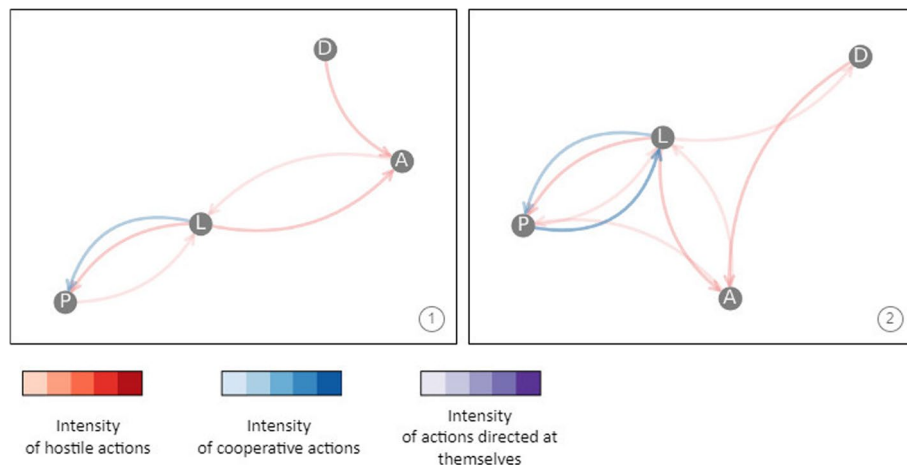


**Fig. 19** Trust-only network for quick season S1. Notes: The number on the arrows indicates the number of trust interactions in each direction

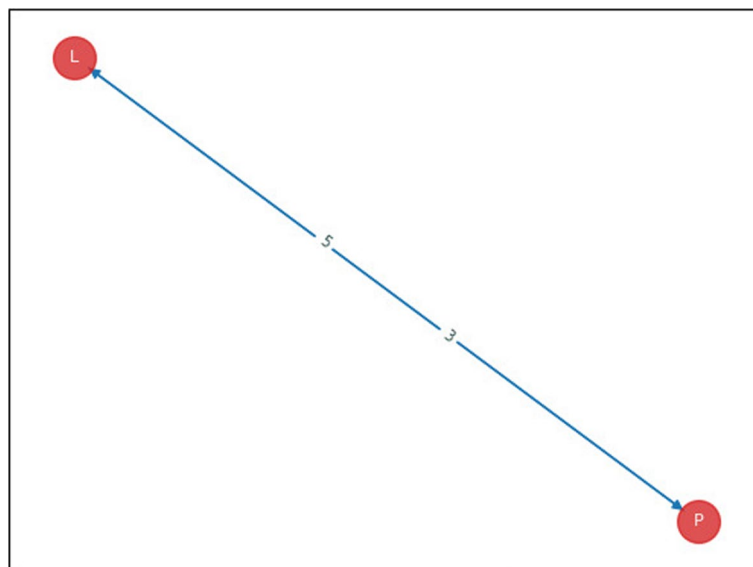


**Fig. 20** Salient states of the general network directly after a trust action, numbered in ascending order. Notes: Red: hostile, blue: friendly, purple: neutral/self-directed. Each node, including the letter, represents a game character

When mutual trust occurred, these were – on average – significantly more pronounced in intensity than the unilateral trust relationships from which they arose. In the process of eliciting mutual trust, some achieved relatively high levels of trusting exchanges with each other, see Table 5. Effects of reciprocity, which are also known from research on non-digital trust, also seem to occur here. A direct comparison of



**Fig. 21** Salient states of the general network directly after a trust action, numbered in ascending order. Notes: Red: hostile, blue: friendly, purple: neutral/self-directed. Each node, including the letter, represents a game character



**Fig. 22** Trust-only network for quick season S3. Notes: The number on the arrows indicates the number of trust interactions in each direction

the prevalence of unilateral and bilateral trust on the Internet and in the analog space would be particularly interesting. An implementation of the presented experimental game in the analog environment could be a further object of research for this purpose.

**Research question 5: the attraction of similar general trust or general distrust**

In addition to the game actions, chat logs, and verbal logs, participants’ general trustfulness was assessed via a questionnaire (SOEP-Trust, see Appendix: *SOEP-Trust*). This score was attributed to the respective nodes to answer Research Question 5:

**Table 5** Sorting unilateral and bilateral (mutual) trust actions across seasons

Season/ Number of Experiment	Unilateral trust	Mutual trust
1	Two Edges with medium intensity (2;3)	One Edge relatively balanced with high intensity (5 and 7)
2	Two Edges with low intensity (1;1)	One Edge balanced with high intensity (5 and 5)
3	Two edges with low intensity (1;1) and another three edges with high intensity (3;4;6)	Two Edges with unbalanced trust (8 and 1; 2 and 6)
4	One edge with medium intensity (2)	One edge with unbalanced trust (11 and 2) and two edges with balanced trust (7 and 8; 3 and 2)
5	Three edges with low/medium intensity (1;2;2) and another edge with high intensity (5)	Three relatively balanced edges, two of them with high intensity (4 and 4; 15 and 11) and another with low intensity (2 and 1)
6	Four edges of middle/high intensity (2;2;3;4)	–
S1	Two edges with low intensity (1;1)	One edge relatively balanced (3 and 5)
S2	–	–
S3	–	One edge relatively balanced (3 and 5)

*Count of trust action(s) are in parentheses, the semicolon is the separator between the different edges*

Related to the allocation of trust, do players favor trust building with strangers of similar low or high self-claimed trust behavior?

To determine whether high or low degrees of trust attract each other, the degree assortativity of the graph was tested once for the trust-only network and then again for the entire network.

For a network  $N, N = (V, E)$  with  $V$  as a set of nodes and  $E$  as a set of directed edges,  $P(v)$  represents each node’s  $v$  characteristic, which is the trust index in this case. In and/or out degrees are indexed by  $\alpha, \beta$ , and  $e_{\{i,j\}}$  is the proportion of directed edges, with  $\alpha$ -degree ( $u$ ) =  $P_i$  and  $\beta$ -degree ( $v$ ) =  $P_j$ , and standard deviations  $\sigma$ . With this, the degree assortativity coefficient can be determined by  $r(\alpha, \beta) = \frac{\sum_{i,j} P_i P_j (e_{ij} - a_i b_j)}{\sigma_a \sigma_b}$ . (Foster, et al. 2010; Newman 2003; Barrat, et al. 2004).

For the network in which only the trust relationships were modeled, an assortativity coefficient of 0.944, i.e. a very high homophily, could be determined. For the entire network, which only asked whether any interaction between two actors had taken place, this was only 0.202, which provides an indication for strong homophily regarding the own statements in terms of the willingness to trust. Players who had low or high trust scores were more likely to form trust relationships with each other than players with widely varying trust scores. According to this finding, homophily can be stated: Individuals with similar trust attitudes seem to be more likely to form edges.

**Discussion**

This study aimed to show trust bonding behavior and effectiveness in an anonymous online environment. Over the course of the nine seasons in total, a multitude of interesting patterns, differing behaviors and socially informed game dynamics manifested. These are complex scenarios that should not be oversimplified with a monodimensional lense and may offer a number of opportunities for further research.

Contrary to initial expectations of pure hostility dominating the gameplay, the findings revealed a nuanced landscape of interactions. While hostile actions were prevalent, instances of trusting interactions emerged between players as well. This observation underscores the complexity of strategic decision-making, where players adapt their approach beyond mere aggression. Subsequently, the outcomes of the nine iterations are contextualized and deliberated upon, in order to facilitate a more profound understanding.

#### 1. Mutual trust and alliance formation:

The theme of mutual trust and alliance formation has been a focal point across various runs. In Season 1, alliances were formed in response to dominant players' actions, showcasing players' adaptation to power dynamics. In Season 3's normal-paced game, an alliance of P, D, T, and F exemplified the collaborative strategies that can emerge. This alliance's trust-based dynamics culminated in F's shared victory, highlighting the reciprocity inherent in alliances. Meanwhile, in the accelerated context of Season S1, the alliance of L and P demonstrated the rapid adaptation and negotiation that characterize time-sensitive gameplay. By situating these findings within the broader context of established research on mutual trust and alliance dynamics, this analysis contributes to the ongoing discourse on the intricate interplay of social phenomena, strategic decision-making, and cooperative behaviors in competitive settings. In the realm of social psychology and game theory, the observed patterns of mutual trust and alliance formation within the competitive gameplay find resonance in established literature. Notable concepts such as Axelrod's 'The Evolution of Cooperation' (1984) and Ostrom's research on trust and reciprocity mechanisms (2000) shed light on the emergence of cooperative behaviors and the establishment of trust-based relationships in scenarios involving self-interested individuals (Axelrod 1984; Ostrom 2003). Also, Nowak and Sigmund's investigation into indirect reciprocity offers theoretical parallels to the mutual trust and alliance dynamics observed in this study (Nowak and Sigmund 2005). These research on cooperation evolution through various mechanisms echoes the emergence of strategic alliances in the competitive gameplay, showcasing how trust and cooperation can emerge and prevail, at least temporarily.

#### 2. Betrayal and deception:

The phenomenon of betrayal and deception was consistently present in the gameplay. From early-game eliminations in Season 1 to strategic misdirections in Season 2, players utilized betrayal as a means to secure their positions. This theme was pronounced in Season Q2, where D's manipulation and proxy wars underscored the strategic complexity of the accelerated environment. Similarly, in Season 8, the absence of trust interactions suggests a heightened urgency that may discourage the establishment of reciprocal relationships. In the realm of social science and social psychology research, the phenomena of betrayal and deception have been subject to investigation (Jones, et al. 1997; Hyman 1989). The utilization of betrayal and deception as deliberate strategies has been recognized for an extended period, traceable in part to the early

writings of Sun Tzu, as well as references by Aristotle and Machiavelli (Wanasika and Adler 2011). Personal motives, social norms, and situational factors exert influence on their manifestation (Kamila, et al. 2012). Moreover, the consequences examined in scientific inquiry, such as negative emotions and profound implications for trust and willingness to cooperate, especially in relationships that have not been ongoing for an extended period (Levine, et al. 2010), have been corroborated within the context of games.

### 3. Resource distribution and cooperation:

Resource distribution and cooperation emerged in multiple runs. In Season 2, D and F's collaboration exemplified the potential of coordinated resource allocation. In Season 4, players D and F's strategy of supporting one another's attacks underscored the power of resource pooling. Additionally, the transfer of resources from G to D in Season S1 highlighted the strategic negotiation that can shape alliances. The transfer of resources is considered a central indicator of cooperation or trust in game theory research (Camerer 2003; Neumann and Morgenstern, 1944). In the present game, trust transfers were not necessary but remained possible. What makes this particularly intriguing is that they were frequently utilized and persisted across multiple players through chains of trust, signifying the successful continuation of trust transfers among (almost) strangers. Similar effects are observed in economic phenomena along their value chains, where multiple actors are involved, but typically rely on long-term positive relationships or contracts to function effectively (Sahay 2003). Therefore, the finding is noteworthy as it demonstrates that strangers, who had every reason to distrust each other and were unable to establish secure contracts, resorted to similar patterns of behavior.

### 4. Influence of power dynamics:

Power dynamics and influence consistently shaped gameplay interactions. Dominant players exerted influence on alliances and decisions. In Season 1, influential players dictated alliance dynamics, while in Season 6, continuous online presence empowered D and G's dominance. Notably, the last-minute elimination of D by G in Season S1 showcased the manipulation and psychological warfare that arise from power dynamics. This type of power dynamics was expected in this type of game, and similar structures can be found in other social deception games. In a large-scale investigation of a negotiation-based game called *Diplomacy*, a close relationship was found between the maintenance and dynamics of power and the deception strategies that follow such dynamics (Denis Peskov 2020). A very similar interdependence can also be presumed for the current design.

### 5. Impact of time constraints

The impact of time constraints introduced by fast quick play rules was implemented in the final three runs. Season S3 demonstrated the heightened urgency of decision-making, influencing the dynamics of alliances, betrayals, and strategic maneuvering. Season S1 highlighted how rapid gameplay fostered swift decision-making, while Season S2

underscored the absence of trust dynamics within an accelerated context. In the realm of social engineering, time pressure is frequently employed (Hadnagy 2010). Research has demonstrated that time pressure can increase the likelihood of individuals falling victim to social engineering attacks, as they tend to be less attentive, less critical, and less rational in their responses (Chowdhury, et al., 2018)s. However, within the rapid-play iterations conducted, a definitive answer to whether time pressure indeed leads to an increased likelihood of expeditious trust granting could not be ascertained. In at least one of the iterations, time pressure resulted in no trust transfers, suggesting that the underlying mechanism may be more complex than a simple equation where time pressure invariably heightens trust allocation.

Based on these discussions, the multifaceted nature of the social phenomena emerging within this competitive game should have become evident. Furthermore, these phenomena are now poised to be contextualized in relation to the focal phenomenon under investigation, that of trust. Trust is inherently intertwined with the aforementioned social phenomena and dynamic processes, existing in a symbiotic relationship that underscores its intricate role within the framework of this study.

Besides the occurrence of trust situations ('trust-as-choice'), general trust towards strangers ('trust-as-attitude') was measured in order to get a holistic overview of trusting behavior (Li 2007, 2008). Here it could be shown that similar attitudes attracted each other when it came to establishing trust bonds that were also reflected in behavior. Homophily seems to be of some importance in the formation of trust bonds and their robustness, as greater homophily appears to further bonding.

Trust played a pivotal role in establishing a sense of perceived safety within a new game season featuring unfamiliar players. This initial reliance on trust facilitated the charting of player intentions, potential hostilities, and alliances aimed at preempting potential threats. Furthermore, alliances formed around trust were employed to enhance resource-sharing strategies, thereby bolstering field control and improving the odds of survival and ultimate victory. Intriguingly, trust was also harnessed as a tool for instilling fear or exerting influence over other players to perform specific actions, as observed in Season S1. This utilization of trust transcended mere collaboration, transforming it into a mechanism for strategic field control and even suppression through tactical alliances. Notably, such maneuvers occurred notably during the revival of players, only to be subsequently eliminated again through alternative alliances, serving as a means to assert dominance on the playing field.

Moreover, trust emerged as a calculated means to secure victory through subterfuge and incentivizing players to engage in targeted hostile actions against their peers. A compelling instance of this occurred when a player offered and followed through with a real-money bribe surpassing the prize money for winning the game. This unique scenario showcased bilateral trust, where one player trusted another to execute actions to ensure their ally's victory while simultaneously accepting their own defeat. This phenomenon underscores the remarkable strength of trust bonds that can form in anonymous online environments. Additionally, a pattern of reliable reciprocal behavior was consistently observed. This pattern may raise questions about the conditions under which this behavior emerges reliably in online settings and the driving forces behind it. Further exploration through the analysis of verbal protocols might shed light on players'



underlying motivations. It is evident that the decision to trust certain players is not necessarily a conscious choice but rather a foundational precondition for the manifestation of overt trust behaviors and the establishment of trust bonds.

It is apparent that the decision to trust certain players is not necessarily made consciously but a mandatory precondition for the emergence of manifest trust behavior and the formation of trust bonds, as has been stated in regards of more general online social networks (ONS) (Grabner-Kräuter 2009).

Generally, it can be constated that there is an air of careful foreignness in the beginning of each game that, sooner or later, may result into (trust) relationships and some familiarity between players.

Embeddedness as such, as well as relational and structural, has been speculated to be influential on the members of a social network system (Granovetter 1992). The results of this study seem to confirm this approach, as game survivability and success seem to be severely tied to positive embeddedness (or skillful deception) in most runs.

Trust behavior mostly manifested between dyads or triads in a meaningful, bilateral way. Trust networks with more nodes, especially those that are of a chainlike structure, may not share the same robustness and contain a number of unilaterally committed satellite players that may attempt to attach to a central figure. This may be in terms of charisma, tactical value or both.

These findings might be applicable to other (mostly) anonymous online activities and trust-bonding in their framework, such as social media in general or specifically chat rooms, online gaming, and others.

While the relative easiness with which trust bonds could be formed in a scenario of mutual suspicion can be interpreted as a hopeful means to overcome initial foreignness in online framework, it might also be viewed as a worrisome opportunity for scamming if the trust is secretly unilateral. Certain techniques to convince others to trust could be reproduced in the experiment, as are also used in real-world social engineering. A detailed analysis of the verbalized thought protocols could offer actionable implications to protect potential victims.

In further research, the design could also be repeated in analog space, as many of the regularities distilled seem very similar to the regularities already known. A major advantage of the design—in contrast to known templates—would be the inclusion of hostile environmental conditions and group dynamic processes that do not isolate trust from other essential social processes, yet still render it measurable.

### **Limitations**

Besides the many dynamics, some of them very individual, that were evident in the process of personal trust building in a hostile environment, some limitations should also be mentioned. At first, the unusually high dropout rate in the games with quick rules, in which action point drop was fastened from two a day (standard) to one an hour (quick), was striking. While the players assigned to the standard games hardly withdrew from the study, dropout rates of up to 60% occurred in the quick rules. Apart from that, the times that the games lasted also differed enormously between the runs, and that without including the quick rules. This can be explained by the unique dynamics between the players, which have developed differently in each run. At the same time, however, this

also creates an opportunity to observe different underlying changes of the emergence of trust.

While verbal protocols were taken to contextualize player decisions in-game, these have not been interpreted qualitatively to examine for player motivations, reactions and emotional as well as practical classifications. Detailed qualitative analysis of the verbal protocols might yield (further) insights in mechanisms of bonding and betrayal in an anonymous online environment. Due to the amount of data, this could form the basis of a separate, more in-depth research on trust in a hostile online environment.

As the sample pool of players was taken from a popular internet forum (Reddit) it may be argued that the results of this paper cannot be generalized to a wider population of non-forum affine users of the Internet, who may be overall more distrusting or naïve towards strangers on the internet. In addition, similar to much research in which volunteers are recruited, a self-selection bias is evident. This concerns in particular the runs under implementation of the quick play rules, as already addressed before.

Lastly, it should be noted, that it remains unclear so far how trust building in an online environment fosters a seemingly robust, (game-)lasting bond, as could be seen in some seasons, compared to a loose bilateral relationship that results in opportunistic 'back-stabs'. Personal preferences and opportunity may be at play here.

## Conclusion

In conclusion, the conducted experiment centered around investigating trust development among initially unfamiliar individuals in an online context. The findings of this study shed light on the intricate dynamics of trust formation within digitally mediated interactions, demonstrating how individuals can navigate uncertainty and risk to establish trusting relationships even in virtual, unreliable (considering the participants' intentions) environments.

The implications of these results extend beyond the experimental setting and may have relevance for real-world scenarios, such as the realm of social engineering. The observed mechanisms of trust-building in an online game highlight the potential vulnerability of individuals to manipulation and deceit, as similar strategies could be exploited by malicious actors in various online and offline contexts. In the context of real-world social engineering, an additional aspect is that there are no established rules of the game, causing the involved individuals to be more careless or unsuspecting. Understanding the nuances of how trust evolves (and gets undermined again) among strangers in a digital environment provides valuable insights into the ways in which trust can be cultivated, manipulated, or compromised.

However, as technological advancements continue to shape interpersonal interactions and redefine the boundaries of trust, the findings of this study underscore that individuals are willing, albeit occasionally, to extend trust even when aware of conditions that would conventionally preclude it. This revelation serves as a testament to the intricate interplay of human decision-making, where individuals, driven by nuanced motivations, may choose to embrace trust despite prevailing skepticism. This can lead to corresponding issues, especially when considering that social engineering is one of the primary attack vectors in the digital realm.

In sum, the study not only illuminates the dynamics of trust formation in seemingly untrustworthy circumstances but also highlights the adaptability of human trust-related behaviors and group dynamics. It emphasizes the need to recognize the multifaceted nature of trust and its potential to transcend conventional barriers, thereby offering a more comprehensive understanding of the complexities that underlie interpersonal relationships in a rapidly evolving technological landscape.

## Appendix

### Game rules

Adapted from Halfbrick Studio's Prototype 'Tank Turn Tactics'

Survive to the end and win the cash.

1st \$50 | 2nd \$25 | First Blood: \$5.

### Rules

- All players start at a random location on the grid, and have 3 hearts and 1 Action Points.
- Every 24 h on a work day, everyone will receive 1 Action Point (AP). The time the point is given may vary day to day. When *Quick Rules* are applied, everyone will receive 1 AP every hour.
- At any time you like, you can do one of the four following actions:
  1. Move to an adjacent, unoccupied square (1 AP)
  2. Shoot someone who is within your range (1 AP). Shooting someone removes 1 heart from their health.
  3. Add a heart (3 AP)
  4. Upgrade your range (3 AP)
- At the start of the game, everyone has a range of 2. That is, they can shoot or trade with someone within 2 squares of them. Upgrading your shooting range increases this by 1 square each time.
- If a player is reduced to 0 hearts, then they are dead. Any action points the dead player had are transferred to the player who killed them. Dead players remain on the board and not removed.
- Players are able to send gifts of hearts or actions points to any player currently within their range.
- Dead players can have a heart sent to them. This will revive that player who will have 1 heart and 0 AP.

**Additional notes**

- Dead players form a jury. Each day they vote, and whoever received most votes will be 'haunted', and not revive any AP for that day.
- Once a day, at a random time, a heart will spawn on the field. The first player to move into the square containing the heart will revive an additional heart.
- The game ends when a clear 1st and 2nd place can be determined.
- Action points are secret! Probably a good idea to try and hide how many you have.
- You can't win this game without making some friends and stabbing some backs. Probably.

**soep-trust**

In general, you can trust people.

- disagree strongly
- disagree somewhat
- agree somewhat
- agree strongly

Nowadays, you can't rely on anybody.

- Disagree strongly
- Disagree somewhat
- Agree somewhat
- Agree strongly

How much do you trust strangers you meet for the first time?

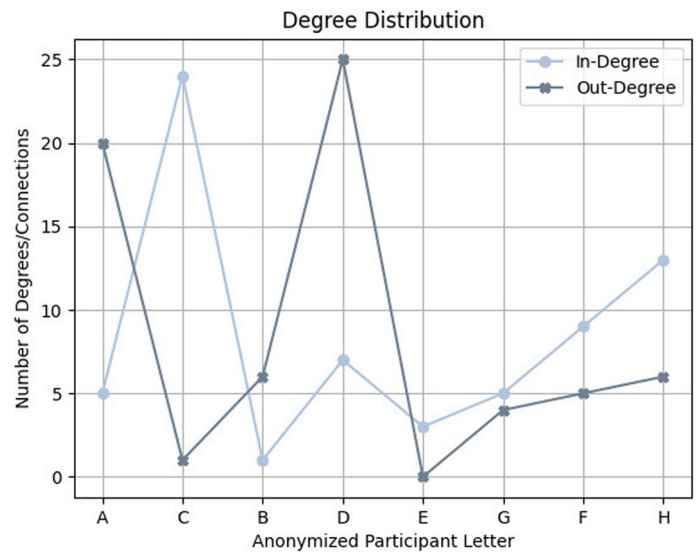
- no trust at all
- little trust
- quite a bit of trust
- a lot of trust

When dealing with strangers, it's better to be cautious before trusting them.

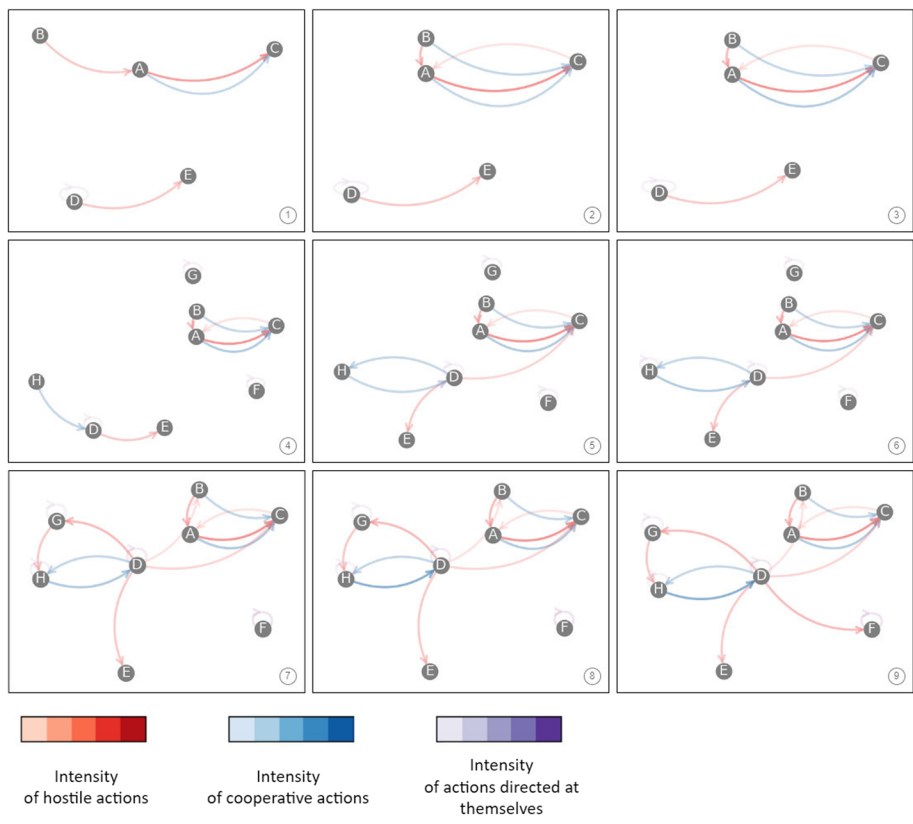
- disagree strongly
- disagree somewhat
- agree somewhat
- agree strongly

**Degree distributions and main networks**

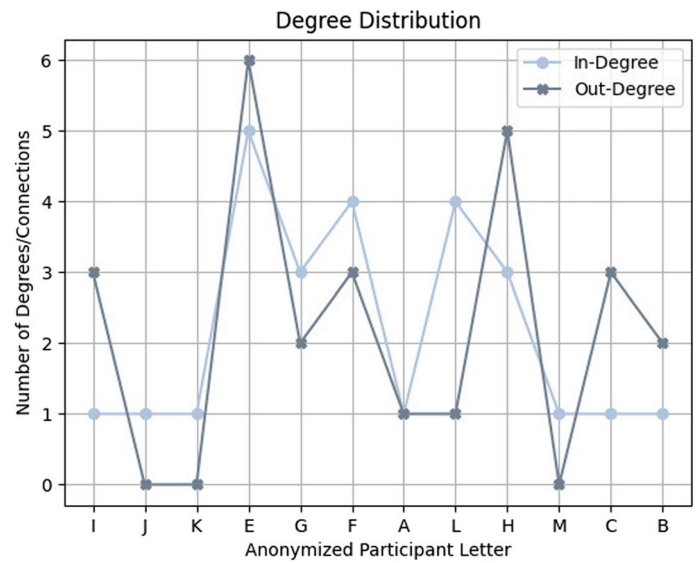
See Figs. 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43 and 44.



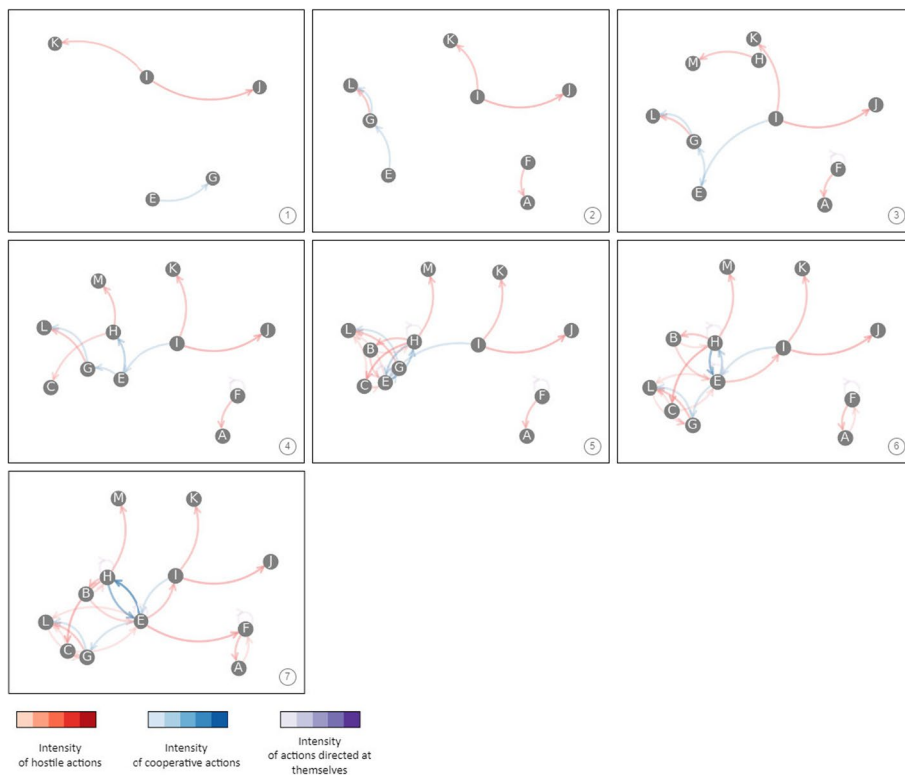
**Fig. 23** Season 1, degree distribution



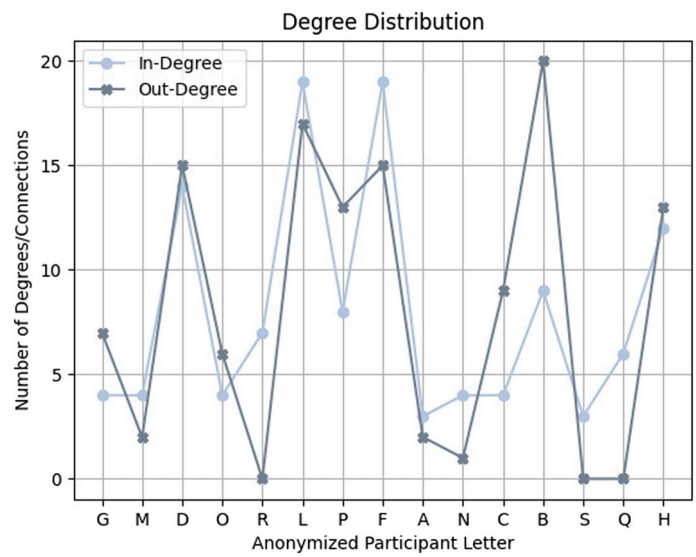
**Fig. 24** Season 1, main network



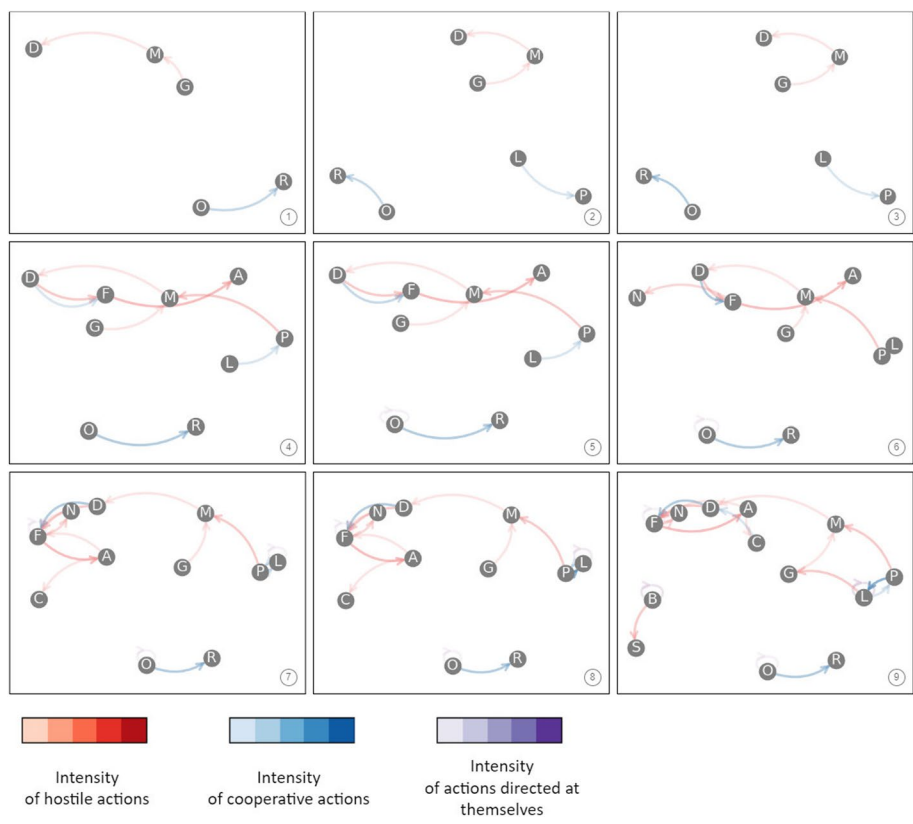
**Fig. 25** Season 2, degree distribution



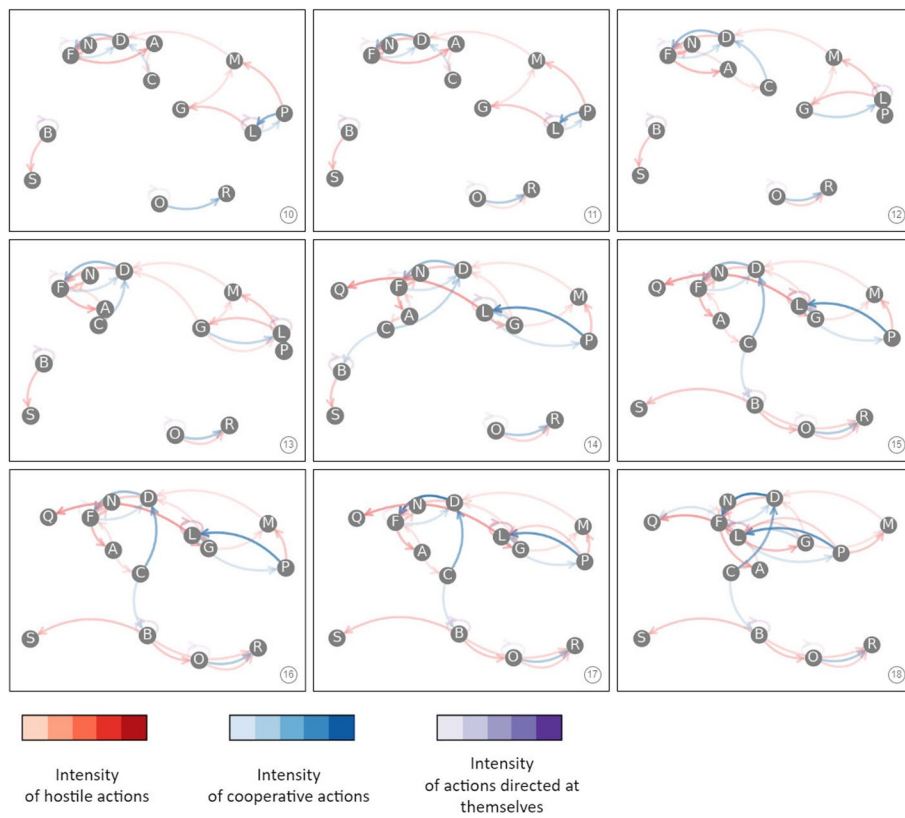
**Fig. 26** Season 2, main network



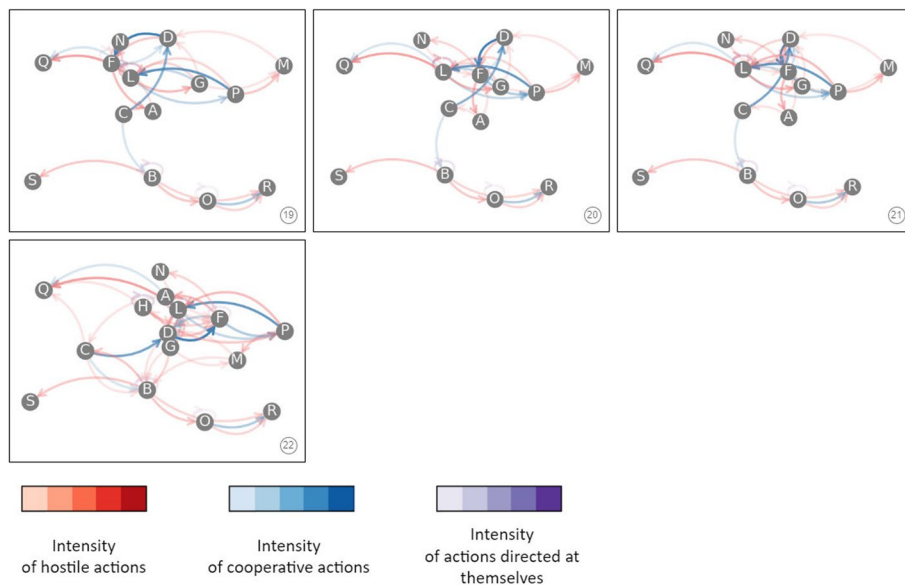
**Fig. 27** Season 3, degree distribution



**Fig. 28** Season 3, main network 1

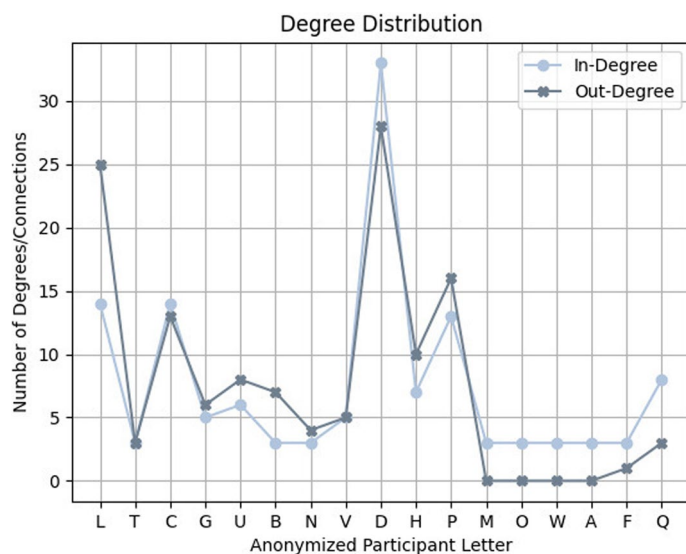


**Fig. 29** Season 3, main network 2

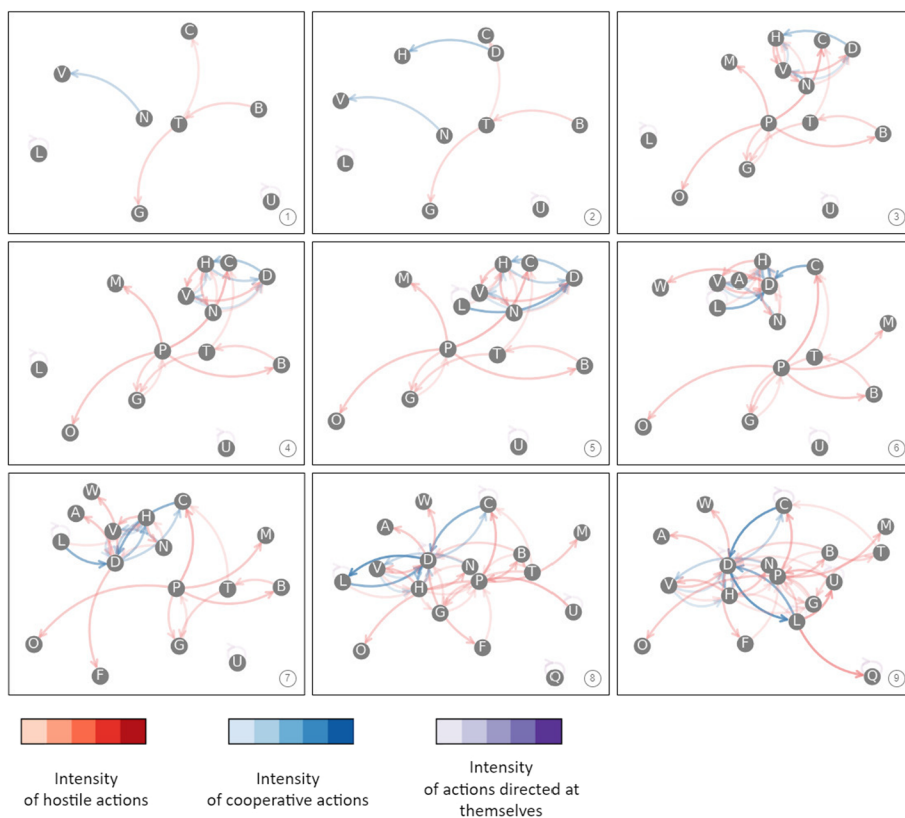


**Fig. 30** Season 3, main network 3

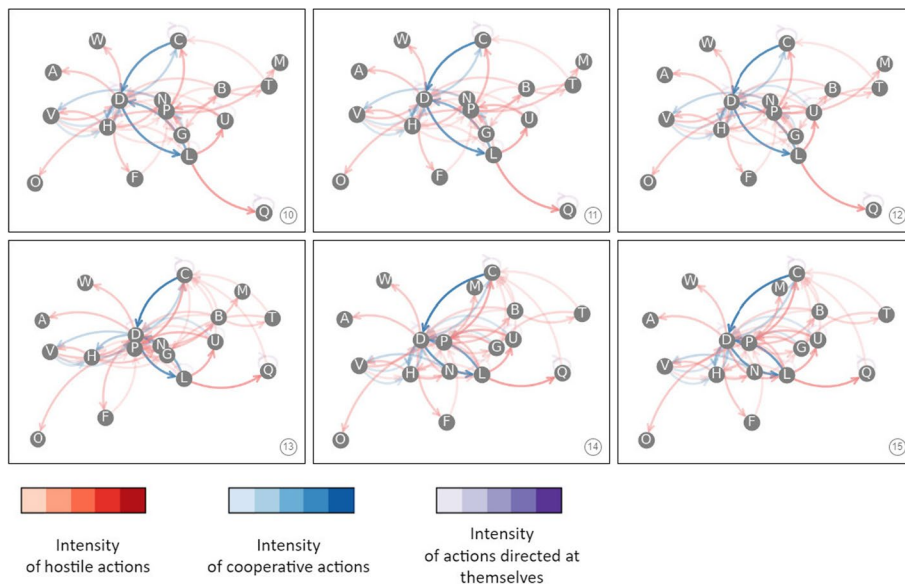




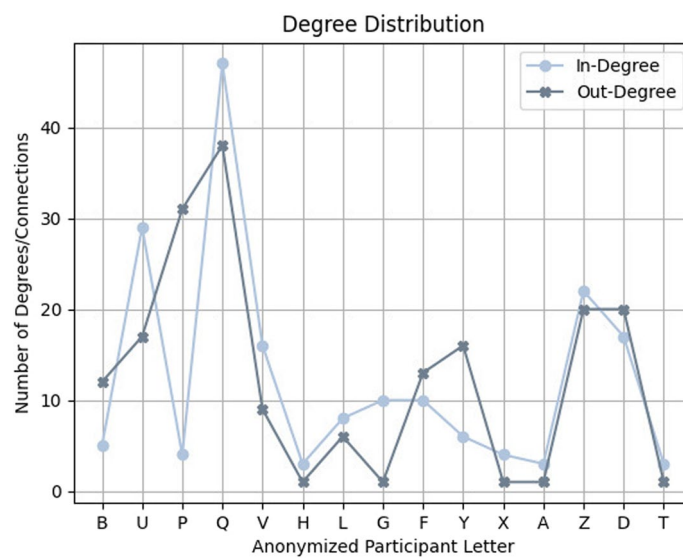
**Fig. 31** Season 4, degree distribution



**Fig. 32** Season 4, main network 1

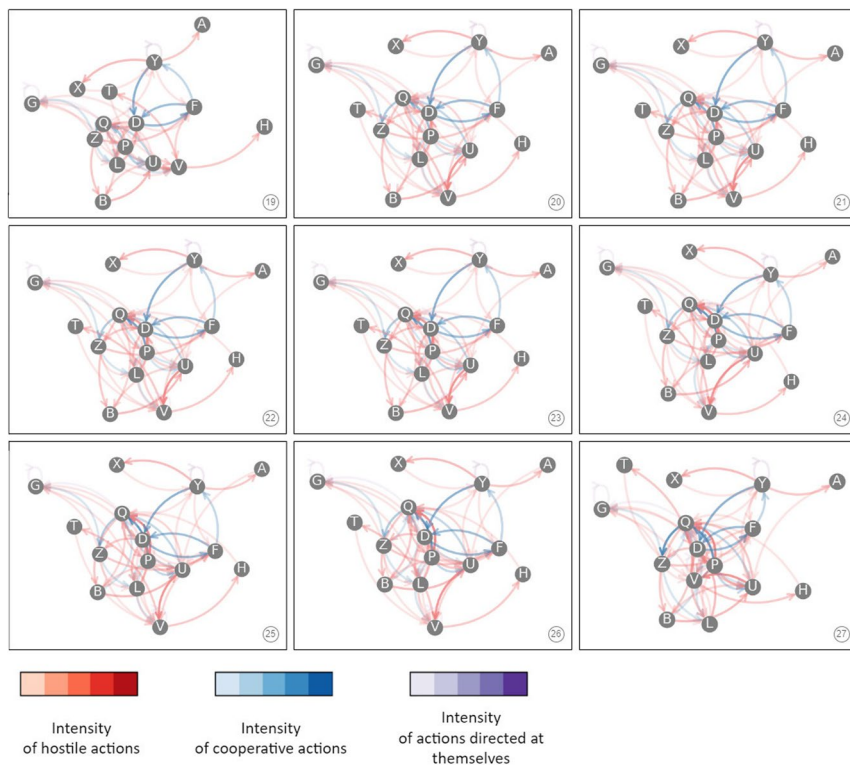


**Fig. 33** Season 4, main network 2

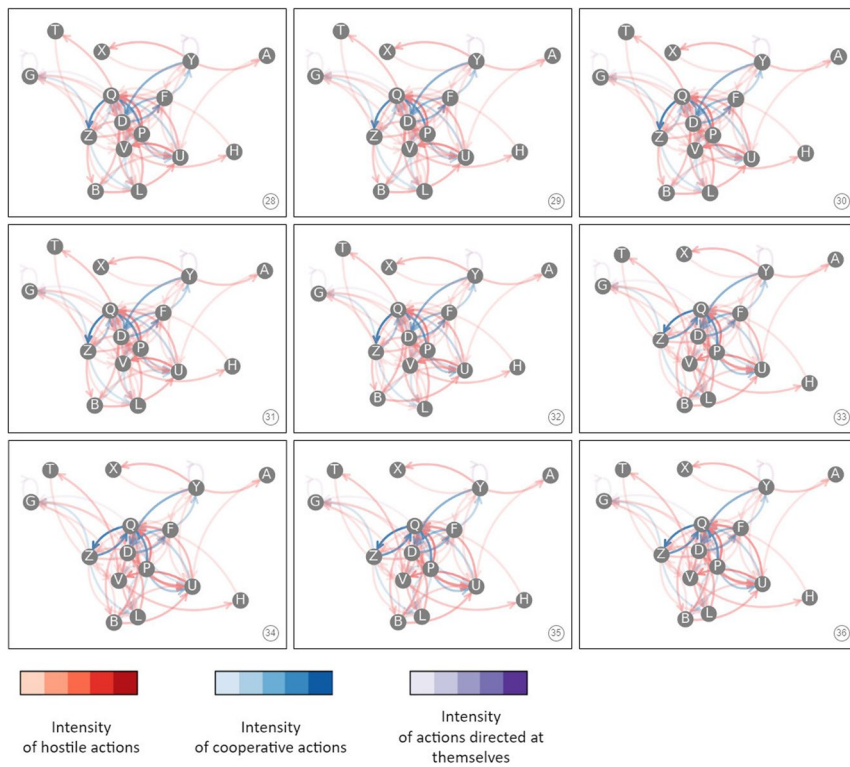


**Fig. 34** Season 5, degree distribution

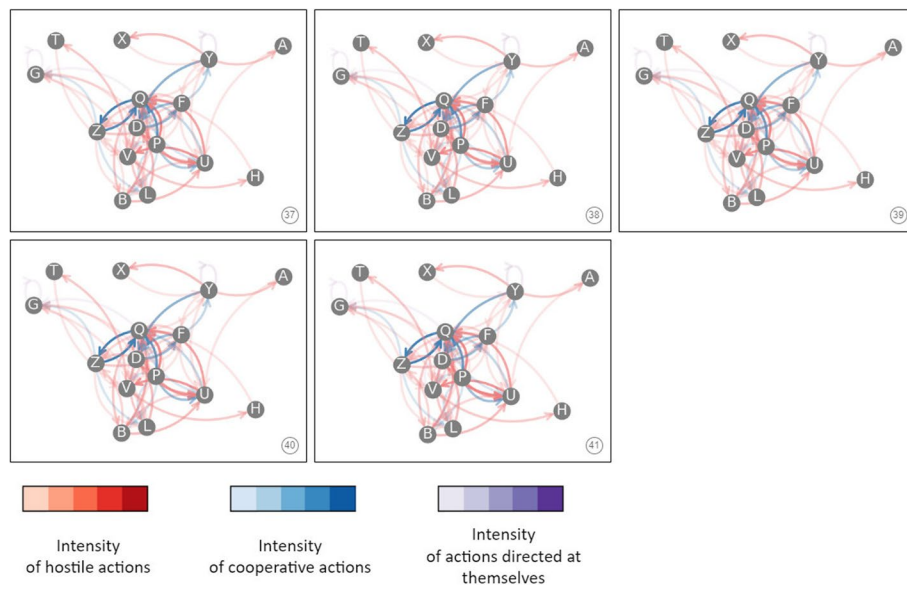




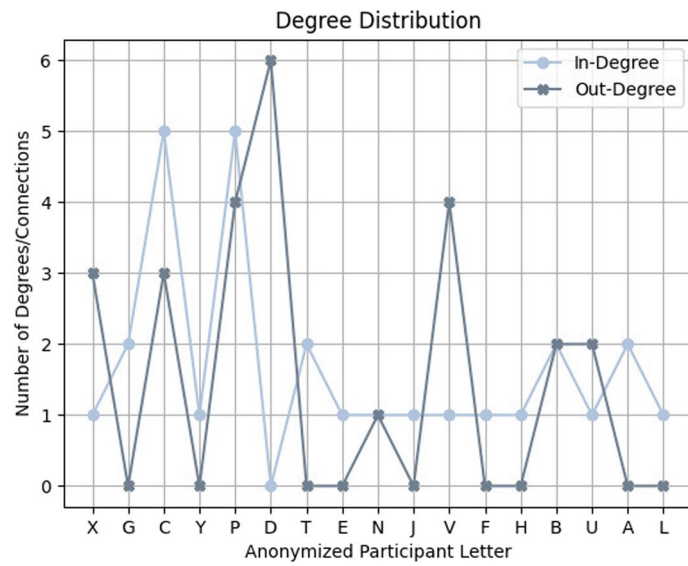
**Fig. 37** Season 5, main network 3



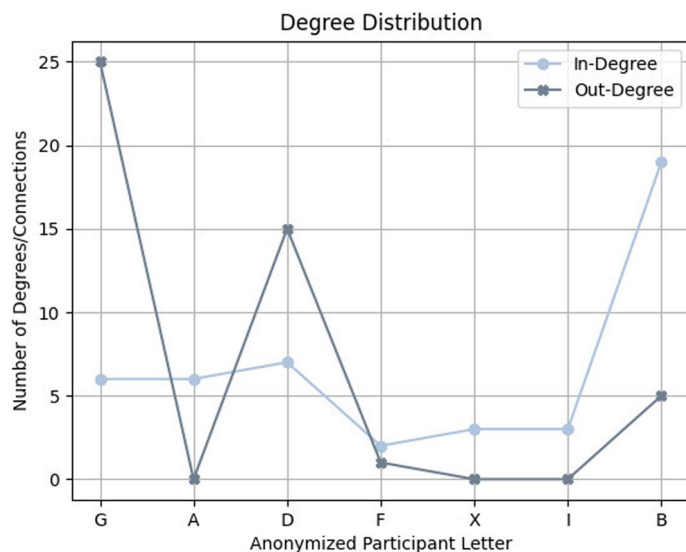
**Fig. 38** Season 5, main network 4



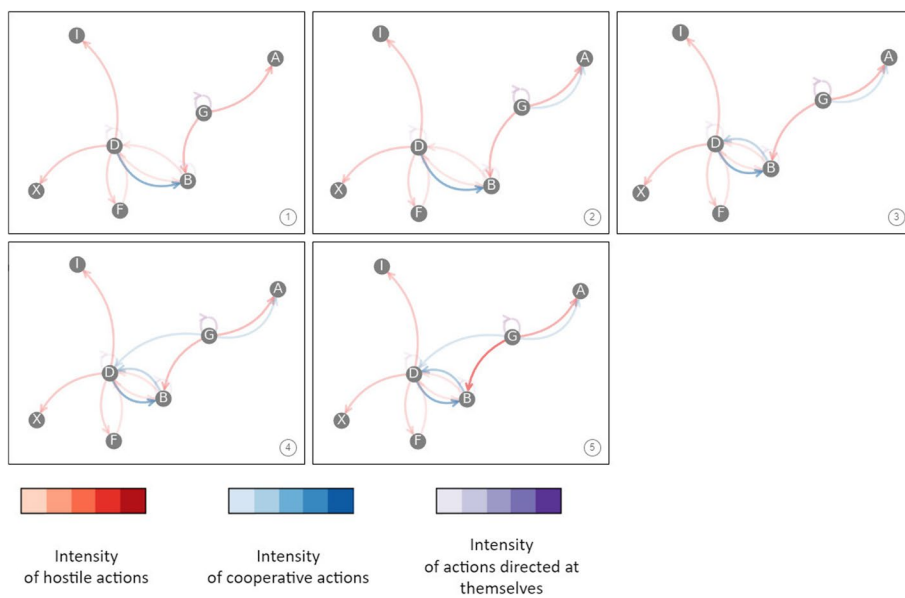
**Fig. 39** Season 5, main network 5



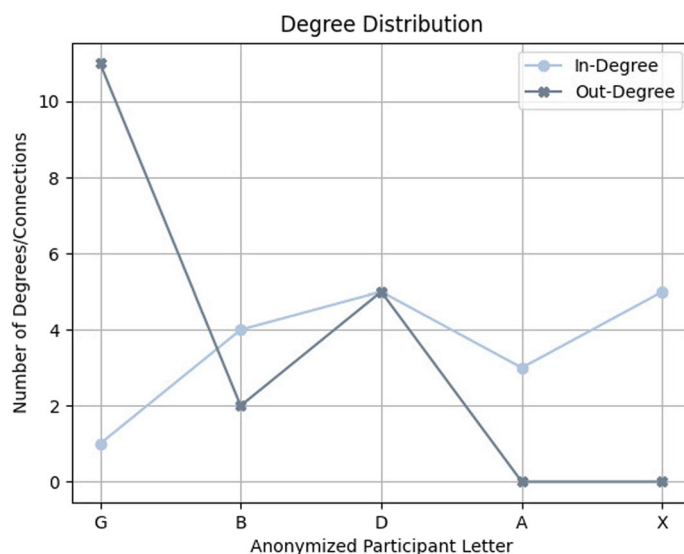
**Fig. 40** Season 6, degree distribution



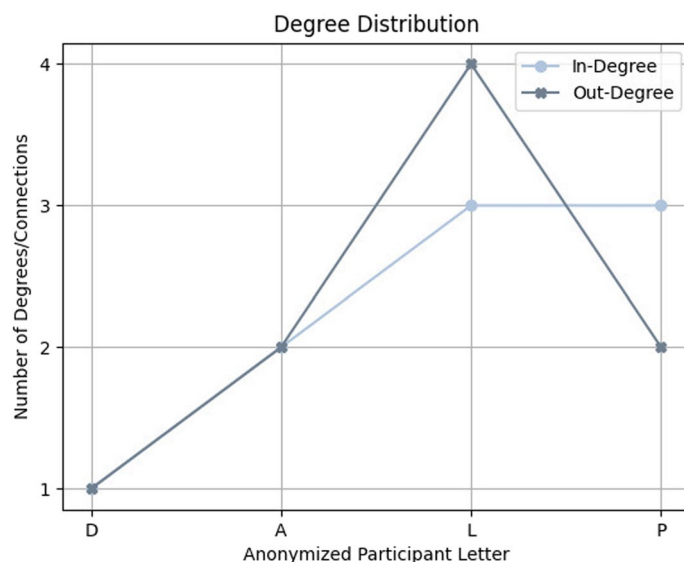
**Fig. 41** Quick Rules 1, degree distribution



**Fig. 42** Quick Rules 1, main network



**Fig. 43** Quick Rules 2, degree distribution



**Fig. 44** Quick Rules 3, degree distribution

**Abbreviations**

- MMSNA Mixed Methods Social Network Analysis
- ONS Online Social Networks
- SNA Social Network Analysis
- SOEP-Trust Socio-Economic Panel Trust in Stranger scale
- WVS World Value Survey

**Acknowledgements**

Not applicable.

**Permission to reproduce material from other sources**

All necessary permissions to reproduce materials from other sources have been obtained.

**Author contributions**

Study Design, Material Preparation, Data Collection and Data Analysis: ALF Editing and Proof Reading: UE All authors read and approved the final manuscript.

**Funding**

Open Access funding enabled and organized by Projekt DEAL.

**Availability of data and materials**

The raw data generated and the data's preparation analysed during the current study are available from the corresponding author on reasonable request.

**Declarations****Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

Received: 19 October 2023 Accepted: 8 March 2024

Published online: 18 March 2024

**References**

- Abraham B, Parigi P, Gupta A, Cook KS (2017) Reputation offsets trust judgments based on social biases among Airbnb users. *Proc National Acad Sci* 114(37):9848–9853
- Alós-Ferrer C, Farolfi F (2019) Trust Games and Beyond. *Front Neurosci* 13:469238
- Andrei F, Barrera D, Krakowski K, Sulis E (2023) Trust intermediary in a cryptomarket for illegal drugs. *Eur Sociol Rev* 40(1):160–172
- Asim Y, Malik AK, Raza B, Shahid AR (2019) A trust model for analysis of trust, influence and their relationship in social network communities. *Telematics Inform* 36:94–116
- Axelrod, R., 1984. *The Evolution of Cooperation*. New York: s.n
- Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. *PNAS* 101(11):3747–3752
- Berg J, Dickhaut J, McCabe K (1995) Trust, Reciprocity, and Social History. *Games Econom Behav* 10(1):122–142
- Boon S, Holmes J (1991) The dynamics of interpersonal trust: Resolving uncertainty in the face of risk. *Cooperation and prosocial behaviour*. Cambridge University Press, Cambridge, pp 190–211
- Bouncken RB, Gast J, Kraus S, Bogers M (2015) Coopetition: a systematic review, synthesis, and future research directions. *RMS* 9:577–601
- Buskens V (2002) *Social Networks and Trust*. Springer Science & Business Media, Luxemburg
- Camerer CF (2003) Behavioral game theory: Experiments in strategic interaction. Princeton Univers. Press, Princeton
- Choi, H. S. & Leon, S., 2023. When trust cues help helpfulness: investigating the effect of trust cues on online review helpfulness using big data survey based on the amazon platform. *Electronic Commerce Research*.
- Chowdhury, N. H., Adam, M. T. P. & Skinner, G., 2018. The Impact of Time Pressure on Human Cybersecurity Behavior: An Integrative Framework. *2018 26th International Conference on Systems Engineering (ICSEng)*.
- Coleman JS (1990) *Foundations of Social Theory*, Cambridge. The Belknap Press, Mass./London
- Corbitt BJ, Thanasankit T, Yi H (2003) Trust and e-commerce: a study of consumer perceptions. *Electron Commer Res Appl* 2(3):203–215
- Dasgupta P (1988) Trust as a Commodity. *Trust: Making and Breaking Cooperative Relations*. Blackwell, Oxford, pp 49–72
- Dawes RM (1980) Social dilemmas. *Annu Rev Psychol* 31:169–193
- Denis Peskov, B. C., 2020. It Takes Two to Lie: One to Lie, and One to Listen. *Proceedings of ACL*, January.
- Dirks KT, Ferrin DL (2002) Trust in leadership: Meta-analytic findings and implications for research and practice. *J Appl Psychol* 87:611–628
- Duradoni M, Paolucci M, Bagnoli F, Guazzini A (2018) Fairness and trust in virtual environments: the effects of reputation. *Future Internet* 10(6):50–65
- Duradoni M, Collodi S, Perfumi SC, Guazzini A (2021) Reviewing stranger on the internet: the role of identifiability through "reputation" in online decision making. *Future Internet* 13(5):110–132
- Ferrin DL, Dirks KT, Shah PP (2003) Many routes toward trust: A social network analysis of the determinants of interpersonal trust. *Academy of Management*, New York, pp C1–C6
- Ferrin DL, Bligh MC, Kohles JC (2007) Can I trust you to trust me? A theory of trust, monitoring, and cooperation in interpersonal and intergroup relationships. *Group Org Manag* 32(4):465–499
- Foster JG, Foster DV, Grassberger P, Paczuski M (2010) Edge direction and the structure of networks. *Proc Natl Acad Sci USA* 107(24):10815–10820
- Froehlich DE, Rehm M, Rienties BC (2020a) *Mixed Methods Social Network Analysis*. Routledge, London
- Froehlich DE, Van Waes S, Schäfer H (2020b) Linking quantitative and qualitative network approaches: a review of mixed methods social network analysis in education research. *Rev Res Educ* 44(1):244–268
- Fruchterman T, Reingold E (1991) Graph drawing by force-directed placement. *Software-Practice & Experience* 21(11):1129–1164
- Gambetta D (1988) *Trust: Making and breaking cooperative relations*. Basil Blackwell, Oxford
- Girlea, C., Girju, R. & Amir, E., 2016. Psycholinguistic Features for Deceptive Role Detection in Werewolf. In: *Proceedings of NAACL-HLT 2016*. San Diego, California: Association for Computational Linguistics, p. 417–422.
- Glaeser EL, Laibson DI, Scheinkman JA, Soutter CL (2000) Measuring trust. *Q J Econ* 115(3):811–846



- Gnyawali DR, Madhavan R (2001) Cooperative networks and competitive dynamics: A structural embeddedness perspective. *Acad Manag Rev* 26(3):431–445
- Grabner-Kräuter S (2009) Web 2.0 social networks: the role of trust. *J Bus Ethics* 90(4):505–522
- Granovetter M (1992) Economic Institutions as Social Constructions: A Framework for Analysis. *Acta Sociologica* 35(1):3–11
- Hadnagy C (2010) *Social Engineering: The Art of Human Hacking*. John Wiley & Sons, New Jersey
- Hargittai E, Fullerton L, Mechen-Trevino E, Thomas KY (2010) Trust Online: Young Adults' Evaluation of Web Content. *Int J Commun* 4:468–494
- Hart, R., 2021. *What is RMT, and how does it happen, and is it allowed?*. [Online] Available at: <https://proprivacy.com/blog/rmt-mmo>
- Hieronana AT, Nugraha AKNA (2021) The influence of social factors, trust, website quality, and perceived risk on repurchase intention in e-commerce. *Jurnal Bisnis Dan Manajemen* 8(2):321–335
- Hirschman AO (1970) Exit Voice and Loyalty, Responses to Decline in Firms, Organizations, and States. Harvard University Press, Cambridge
- Ho JC-T (2020) How biased is the sample? Reverse engineering the ranking algorithm of Facebook's Graph application programming interface. *Big Data Soc* 7(1):1–15
- Hyman R (1989) The Psychology of Deception. *Annual Rev Psychol* 40:133–154
- Jack S (2010) Approaches to studying networks: implications and outcomes. *J Bus Ventur* 25(1):120–137
- Jackson MO, Watts A (2002) The evolution of social and economic networks. *J Econ Theory* 106(2):265–295
- Jones WH, Couch L, Scott S (1997) Trust and Betrayal: The Psychology of Getting Along and Getting Ahead. In: Hogan R, Johnson J, Briggs S (eds) *Handbook of Personality Psychology*. Academic Press, London et al., pp 465–482
- Jøsang, A. & Presti, S. L., 2004. Analysing the Relationship between Risk and Trust. In: *International Conference on Trust Management*. Oxford, UK: Conference Proceedings, p. 135–145.
- Kamila S et al (2012) What if i get busted? Deception, choice, and decision-making in social interaction. *Front Neurosci* 6:21893
- Knock D, Yang S (2020) *Social Network Analysis. Quantitative Applications in the Social Sciences*, 3rd edn. SAGE Publications, Thousand Oaks, California
- Koehly LM, Pattison P (2005) Random graph models for social networks: Multiple relations or multiple raters. *Models and Methods in Social Network Analysis*. Cambridge University Press, New York, pp 162–191
- Kolleck N (2013) Social network analysis in innovation research: using a mixed methods approach to analyze social innovations. *Eur J Futures Res* 1(25):9–27
- Komorita SS, Parks CD (1995) Interpersonal relations: Mixed-motive interaction. *Annu Rev Psychol* 46(1):183–207
- Kühne, S. & Zindel, Z., 2020. Using Facebook and Instagram to Recruit Web Survey Participants: A Step-by-Step Guide and Application. *Survey Methods: Insights from the Field, Special issue: 'Advancements in Online and Mobile Survey Methods'*, p. <https://surveyinsights.org/?p=13558>.
- Lacey, D. & Salmon, P. M., 2015. It's dark in there: using systems analysis to investigate trust and engagement in dark web forums. *EPCE 2015: Engineering Psychology and Cognitive Ergonomics*, p. 117–128.
- Levine TR, Kim S-Y, Ferrara M (2010) Social exchange, uncertainty, and communication content. *Hum Commun* 13(4):303–318
- Lewicki RJ, Wiethoff C (2000) Trust, trust development, and trust repair. *The handbook of conflict resolution: Theory and practice*. Jossey-Bass Publishers, San Francisco, pp 86–107
- Lewicki RJ, Tomlinson EC, Gillespie N (2006) Models of interpersonal trust development: theoretical approaches, empirical evidence, and future directions. *J Manag* 32(6):991–1022
- Li P (2007) Towards an interdisciplinary conceptualization of trust: a typological approach. *Manag Organ Rev* 3(3):421–445
- Li P (2008) Toward a geocentric framework of trust: an application to organizational trust. *Manag Organ Rev* 4(3):413–439
- Liu Y, Tang X (2018) The effects of online trust-building mechanisms on trust and repurchase intentions: An empirical study on eBay. *Inf Technol People* 31(3):666–687
- Liu B et al (2019) Large-scale group decision making model based on social network analysis: trust relationship-based conflict detection and elimination. *Eur J Oper Res* 275(2):737–754
- Masclat D, Pénard T (2012) Do reputation feedback systems really improve trust among anonymous traders? An Experimental Study. *Appl Econ* 44(35):4553–4573
- Mayer RC, Davis JH, Schoorman FD (1995) An integrative model of organizational trust. *Acad Manag Rev* 20:709–734
- McAllister DJ (1995) Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Acad Manag J* 38(1):24–59
- Mcknight DH, Chervany NL (2001) Trust and Distrust Definitions: One Bite at a Time. *Trust in Cyber-societies: Integrating the Human and Artificial Perspectives*. Springer, Berlin, Heidelberg, pp 27–54
- Moeller, K., 2023. Trust in Cryptomarkets for Illicit Drugs. In: E. P. Limited, ed. *Digital Transformations of Illicit Drug Markets: Reconfiguration and Continuity*. Bingley: Tzanetakis, M.; South, N., pp. 29–43.
- Naef, M. & Schupp, J., 2009. Measuring Trust: Experiments and Surveys in Contrast and Combination. *SSRN Electronic Journal*, pp. 1–44.
- Neumann J, v. & Morgenstern, O., (1944) *Theory of Games and Economic Behavior*. Princeton University Press, Princeton
- Newman MEJ (2003) Mixing patterns in networks. *Phys Rev E* 67(2):026126
- Norbutas L (2020) Trust on the Dark Web: An analysis of illegal online drug markets. Utrecht University, s.l.
- Nowak MA, Sigmund K (2005) Evolution of indirect reciprocity. *Nature* 437(7063):1291–1298
- Ostrom E (2000) Collective action and the evolution of social norms. *J Econ Perspect* 14(3):137–158
- Ostrom E (2003) Toward a behavioral theory linking trust, reciprocity, and reputation. In: Ostrom E, Walker J (eds) *Trust and reciprocity: Interdisciplinary lessons from experimental research*. Russell Sage Foundation, New York, pp 19–79
- Pattison P (1993) *Algebraic Models for Social Networks*. Cambridge University Press, New York
- Prashanth R, Cleotilde G (2018) Creative persuasion: a study on adversarial behaviors and strategies in phishing attacks. *Front Psychol* 9:1–14

- Przepiorka W, Norbutas L, Corten R (2017) Order without Law: Reputation Promotes Cooperation in a Cryptomarket for Illegal Drugs. *Eur Sociol Rev* 33(6):752–764
- Resnick P, Zeckhauser R (2002) Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system. *The Economics of the Internet and E-commerce*. Emerald Group Publishing, Bingley, West Yorkshire, pp 127–157
- Ross W, LaCroix J (1996) Multiple meanings of trust in negotiation theory and research: a literature review and integrative model. *Int J Conflict Manag* 7:314–360
- Rotter JB (1980) Interpersonal trust, trustworthiness, and gullibility. *Am Psychol* 35(1):1–7
- Rousseau DM, Sitkin SB, Burt RS, Camerer C (1998) Not so different after all: a cross-discipline view of trust. *Acad Manag Review* 23(3):07
- Sahay B (2003) Understanding trust in supply chain relationships. *Indust Manag & Data Syst* 103(8):553–563
- Schelling T (1960) *The Strategy of Conflict*. Harvard Univ. Press, Cambridge, MA
- Scott J, Carrington P (2011) *Handbook of Social Network Analysis*. Sage Publications, London
- Sen F et al (2016) Focal structures analysis: identifying influential sets of individuals in a social network. *Soc Netw Anal Min* 6(1):1–22
- Serva MA, Fuller MA, Mayer RC (2005) The Reciprocal nature of trust: a longitudinal study of interacting teams. *J Organ Behav* 26(6):625–648
- Shafie T (2019) A multigraph approach to social network analysis. *J Soc Struct* 16(1):1–21
- Small ML (2011) How to conduct a mixed methods study: Recent trends in a rapidly growing literature. *Ann Rev Sociol* 37:57–86
- Sztompka P (1999) *Trust: A sociological theory*. Cambridge University Press, Cambridge
- Sztompka P (2006) New perspectives on trust. *Am J Sociol* 112(3):905–919
- Walton RE, McKersie RB (1966) Behavioral dilemmas in mixed-motive decision making. *Behav Sci* 11(5):370–384
- Wanasika I, Adler T (2011) Deception as Strategy: Context and Dynamics. *J Manag Issues* 23(3):364–378
- Wang YD, Emurian HH (2005) An overview of online trust: concepts, elements, and implications. *Comput Hum Behav* 21(1):105–125
- Watters PA (2009) Why do users trust the wrong messages? A behavioural model of phishing. Tacoma, WA, IEEE, pp 1–7
- Williams TA, Shepherd DA (2015) Mix Method Social Network Analysis: Combining Inductive Concept Development, Content Analysis, and Secondary Data for Quantitative Analysis. *Organ Res Methods* 20(2):268–298
- Williams EJ, Beardmore A, Joinson AN (2017) Individual differences in susceptibility to online influence: A theoretical review. *Comput Hum Behav* 72:412–421
- Xiong S, Li W, Mao X, Iida H (2017) *Mafia Game Setting Research Using Game Refinement Measurement*. ACE 2017: Advances in Computer Entertainment Technology. Springer International Publishing, Basel, pp 830–846
- Zhang Z, McGettigan C, Belyk M (2022) Speech timing cues reveal deceptive speech in social deduction board games. *PLoS ONE* 17(2):e0263852

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.