

Multirelational organization of large-scale social networks in an online world

Michael Szell^a, Renaud Lambiotte^{b,c}, and Stefan Thurner^{a,d,1}

^aSection for Science of Complex Systems, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria; ^bInstitute for Mathematical Sciences, Imperial College London, London SW7 2PG, United Kingdom; ^cDepartment of Mathematics, Imperial College London, London SW7 2AZ, United Kingdom; and ^dSanta Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501

Edited by H. Eugene Stanley, Boston University, Boston, MA, and approved June 23, 2010 (received for review March 25, 2010)

The capacity to collect fingerprints of individuals in online media has revolutionized the way researchers explore human society. Social systems can be seen as a nonlinear superposition of a multitude of complex social networks, where nodes represent individuals and links capture a variety of different social relations. Much emphasis has been put on the network topology of social interactions, however, the multidimensional nature of these interactions has largely been ignored, mostly because of lack of data. Here, for the first time, we analyze a complete, multirelational, large social network of a society consisting of the 300,000 odd players of a massive multiplayer online game. We extract networks of six different types of one-to-one interactions between the players. Three of them carry a positive connotation (friendship, communication, trade), three a negative (enmity, armed aggression, punishment). We first analyze these types of networks as separate entities and find that negative interactions differ from positive interactions by their lower reciprocity, weaker clustering, and fatter-tail degree distribution. We then explore how the interdependence of different network types determines the organization of the social system. In particular, we study correlations and overlap between different types of links and demonstrate the tendency of individuals to play different roles in different networks. As a demonstration of the power of the approach, we present the first empirical large-scale verification of the long-standing structural balance theory, by focusing on the specific multiplex network of friendship and enmity relations.

complex networks | multiplex relations | quantitative sociology

Human societies can be regarded as large numbers of locally interacting agents, connected by a broad range of social and economic relationships. These relational ties are highly diverse in nature and can represent, e.g., the feeling a person has for another (friendship, enmity, love), communication, exchange of goods (trade), or behavioral interactions (cooperation or punishment). Each type of relation spans a social network of its own. A systemic understanding of a whole society can only be achieved by understanding these individual networks and how they influence and coconstruct each other. The shape of one network influences the topologies of the others, as networks of one type may act as a constraint, an inhibitor, or a catalyst on networks of another type of relation. For instance, the network of communications poses constraints on the network of friendships, trading networks are usually constrained to positively connoted interactions such as trust, and networks representing hostile actions may serve as a catalyst for the network of punishments. A society is therefore characterized by the superposition of its constitutive socioeconomic networks, all defined on the same set of nodes. This superposition is usually called multiplex, multirelational, multimodal, or multivariate network (see Fig. 1). The study of small-scale multiplex networks has a long tradition in the social sciences (1) and has been applied to areas such as homophily in social networks (2), the effect of combined interactions on an agent's behavior (3), and the nontrivial interrelation between family and business networks (4). Multiplexity is thought to play an important role

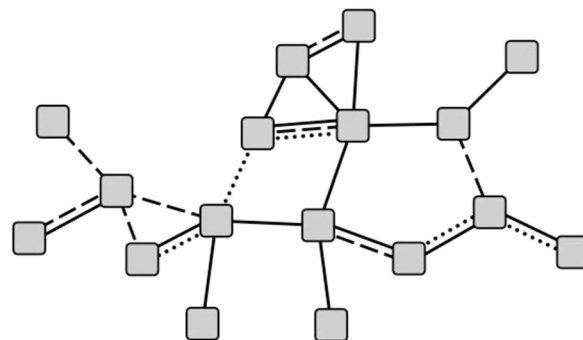


Fig. 1. Multiplex networks consist of a fixed set of nodes connected by different types of links. This multirelational aspect is usually neglected in the analysis of large social networks. In our MMOG dataset, six types of social links can exist between any two players, representing their friendship or enmity relations, their exchanged private messages, their trading activity, their one-to-one aggressive acts against each other (attacks), and their placing of head money (bounties) on other players as, e.g., means of punishment.

in the organization of large-scale networks. For example, the existence of different link types between agents explains the overlap of community structures observed in social networks, where nodes may belong to several communities, each associated to one different type of interaction (5, 6). Methodological work on multiplex networks includes the development of multiplex community detection (7), clustering (8), and other network analysis algorithms (9). The role of multiple relation types in measured social networks has recently been investigated across communication media (10), in an online game (11), as well as in ecological networks (12).

Traditional methods of social science, such as small-scale questionnaire-based approaches, get more and more replaced by automated methods of data collection which allow for entirely different scales of analysis (13–15). This change of scale has opened new perspectives and has the potential to radically transform our understanding of social dynamics and organization (16). The empirical verification of social theories such as the strength of weak ties (17, 18) become possible with hitherto unthinkable levels of precision. However, this large-scale perspective suffers from the drawback of a relatively coarse-grained representation of social processes taking place between individuals and of blindness in respect to the existence of different types of social interactions. For example, in most works on e-mail (19) or mobile phone networks (17, 20), the existence and weight of a link is

Author contributions: M.S., R.L., and S.T. performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹To whom correspondence should be addressed. E-mail: stefan.thurner@meduniwien.ac.at.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1004008107/-DCSupplemental.

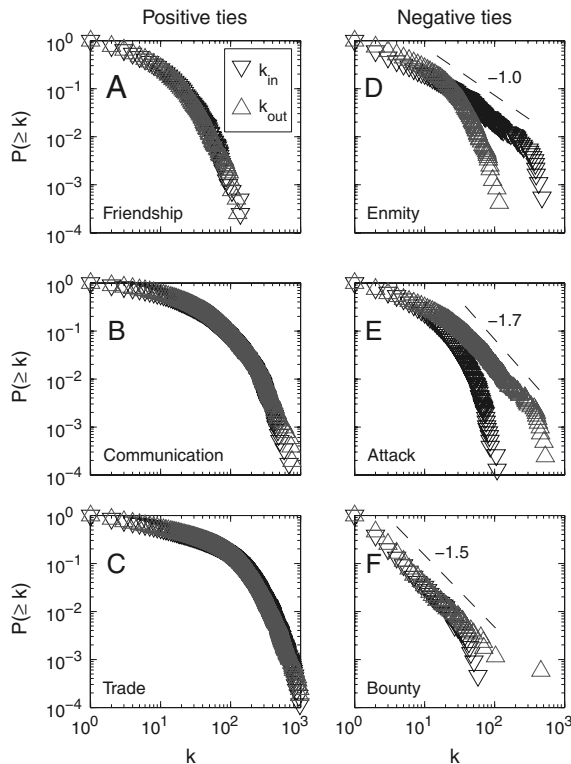


Fig. 2. Cumulative in-degree and out-degree distributions for the six types of networks spanning the same set of agents: (A) friendship, (B) communication, (C) trade, (D) enmity, (E) attack, and (F) bounty. Note the differences between in- and out-degree distributions and the presence of power laws (with cutoffs) for negatively connoted interactions (Right Column), which are absent for positive ties (Left Column). It is immediately clear that topological properties of social networks depend strongly on the nature of their ties. Ignoring this multirelational composition can lead to loss of essential information.

values suggests that mechanisms such as triadic closure (34) are not dominant for negative interactions (see *SI Text* for a confirmation) and has its origin in the balance of signed motifs (see below).

The independent analysis of the different networks reveals distinct types of organization which depend on the nature of the links. It is crucial to account for these distinct topological properties in models for the dynamics of cooperation and conflict in human societies. To demonstrate the danger of not differentiating between types of interactions, we include data on the envelope network (as defined in *Materials and Methods*) in Table 1. Neglecting the nature of social ties and mixing different interactions (even within the same dataset) results in gross misrepresentation of the system, in this case at least by losing the typical low reciprocity and clustering observed in negative tie subnetworks.

For a detailed analysis of the time evolution of single network properties on the same dataset (first 445 days in the Artemis game universe), refer to ref. 18. There several “aging” or “maturing” effects were reported, such as a decrease of the clustering coefficient and reciprocity in friend networks over time.

Network–Network Interactions. Due to strong interactions between different social relations, a next level of complexity enters when considering the coexistence of different types of links (35). From now on, we only focus on undirected versions of the networks, as defined in *Materials and Methods*. To quantify the resulting interdependencies between pairs of networks, we follow two approaches.

On one hand, we focus on the link overlap between networks and calculate the Jaccard coefficient $J_{\alpha\beta}$ between two different sets of links α and β . On the other hand, we compute correlations $\rho(k_\alpha, k_\beta)$ between node degrees in different networks (see *SI Text*). These coefficients measure to which extent degrees of agents in one type of network correlate with degrees of the same agents in another one. If $\rho(k_\alpha, k_\beta)$ is close to one, players who have many (few) links in network α have many (few) links in network β . Note that both measures might be affected by different network sizes or average degrees. To account for this possibility, we additionally compute correlations $\rho(\text{rk}(k_\alpha), \text{rk}(k_\beta))$ between rankings of node degrees, where rk represents rank. Overlap and correlation quantities provide complementary insights into the organization of social structures. In Fig. 3, for all pairs of networks, the three measures are shown. Note that no causal directions can be implied and that all correlations are positive. From highest to smallest overlap (from left to right), Fig. 3 provides the following conclusions:

Communication–friendship. The pronounced overlap implies that friends tend to talk with each other. The equally pronounced correlation attests that players who communicate with many (few) others tend to have many (few) friends. The former result was already reported in ref. 18, where a high fraction of communication partners was shown to be friends.

Trade–communication. The high overlap shows that trade partners have a tendency to communicate with each other, whereas the high correlations shows a tendency of communicators being traders.

Enmity–attack. The high overlap shows that enemies tend to attack each other, or that attacks are likely to lead to enemy markings. The high correlations imply that aggressors or victims of aggression tend to be involved in many enemy relations.

Communication–attack. The relatively high overlap shows that there is a tendency for communication taking place between players who attack each other. The relatively high correlation implies that players who communicate with many (few) others tend to attack or be attacked by many (few) players. Aggression is not anonymous, but accompanied by communication.

Enmity–bounty and attack–bounty. Similar to enmity–attack.

Communication–enmity. Similar to communication–attack.

Trade–friendship. Similar to trade–communication, however with a smaller overlap. It is more difficult for traders to become friends than to just communicate.

Friendship–attack. The low overlap shows that attacks tend to *not* take place between friends, or that fighting players do *not* tend to become friends. The relatively high correlations mean that players with many (few) friends attack or are attacked by many (few) others.

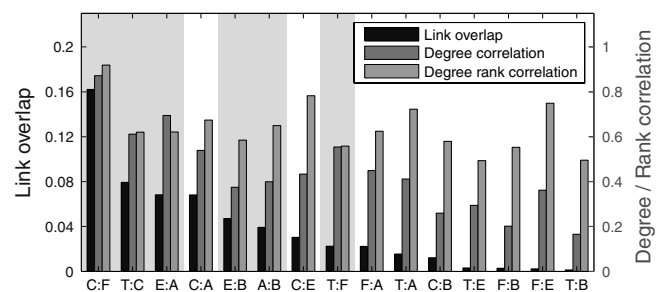


Fig. 3. Link overlap (Jaccard coefficient), degree correlation $\rho(k_\alpha, k_\beta)$, and degree rank correlation $\rho(\text{rk}(k_\alpha), \text{rk}(k_\beta))$ for all pairs of networks (ordered by link overlap). E, enmity; F, friendship; A, attack; T, trade; C, communication; and B, bounty. Pairs of equal connotation (positive–positive or negative–negative) are marked with a gray background. These pairs have high overlaps, whereas oppositely connotated pairs have lower overlaps. The various relations are organized in a nontrivial way, suggesting that agents play very different roles in different relational networks.

Discussion

Most empirical studies of large-scale social networks focus on node properties (5), for instance, to uncover the topological centrality of social agents or patterns of homophily between agents (40), while being blind to the multiple nature of the links connecting agents. In many social systems, however, a proper description of multiplexity is essential to capture the stress caused by different forces acting on social agents and therefore to uncover the principles shaping the large-scale organization of social interactions. For instance, the interaction and coexistence of multiple relations are crucial to describe the emergence of conflict in social systems (41–43) or the development of trust in commercial networks (44).

Our work begins to quantitatively measure the multidimensionality of human relationships. Its results shed light on macroscopic implications of interaction types: Relations driven by aggression lead to markedly different systemic characteristics than relations of a nonaggressive nature. Network–network interactions reveal a nontrivial structure of this multidimensionality and how humans play very different roles in different relational networks. The richness of the dataset allows the effect of multiple relations on the structure and stability of a large-scale social network to be explored, thereby providing a first empirical basis for the modeling of multiplex complex networks. Future research perspectives include different generalizations of structural balance theory, e.g., to a larger set of social relations, to the case of weighted and/or directed networks or to larger motifs, an extension of the concept of modularity for multiplex (7) or signed (45) networks but also dynamical aspects, for instance, the dynamics of noncooperative organizations (46).

Materials and Methods

Social Network Data from the Online Game “Pardus.” The dataset contains practically all actions of all players of the MMOG Pardus since 2004 when the game went online (18). Pardus is an open-ended game with a worldwide player base of more than 300,000 people. Players live in a virtual, futuristic universe in which they explore and where they interact with others in a multitude of ways to achieve their own goals (22). Here we focus on one of the three separate game universes, Artemis, in which $N = 18,819$ players have interacted with at least one other player over the first 445 consecutive days of this universe’s existence.

Players typically engage in various economic activities to accumulate wealth. Communication between any two players can take place directly, by using a one-to-one, e-mail-like, PM system (see *SI Text*), or indirectly, by meeting in built-in chat channels or online forums. Social and economical decisions of players are often strongly influenced and driven by social factors such as friendship, cooperation, and conflict. Conflicting relations may result in aggressive acts such as attacks, fights, revenge, or even destruction of another player’s means of production or transportation. Under certain conditions, hostile acts may degenerate into large-scale conflicts between different factions of players—wars.

The dataset contains longitudinal and relational data allowing for an almost complete and dynamical mapping of multiplex relations of an entire society. The data are free of interviewer effects because agents are not conscious of their actions being logged. Measurement errors which usually affect reliability of survey data (47) are practically absent. The longitudinal aspect of the data allows for the analysis of dynamical aspects such as the emergence and evolution of network structures. Finally, it is possible to extract multiple social relationships between a fixed set of humans. We focus on the following set of six types of one-to-one interactions between players (for details, see *SI Text*): friendship and enmity relations, PM communication, trades, attacks, and revenge/punishment through head money (bounties). We label these networks by Greek indices: $\alpha = 1$ refers to friendship networks, ..., $\alpha = 6$ to bounties. We focus on one-to-one interactions only (without projections as, e.g., used in refs. 48 and 49) and discard indirect interactions such as mere participation in a chat.

Friendship and enmity networks are taken as snapshots at the last available day 445. All other networks are aggregated over time, meaning that whenever a link existed within day 1 and 445, it is counted as a link. For simplicity, we use unweighted, directed networks. Further, we define undirected networks as follows: A link exists between nodes i and j if there exists at least one directional link between those nodes. We construct triads [motifs of three connected nodes (1)] from undirected links. For a combined analysis of the whole system, we define an *envelope network* which is composed of the set of all links of *all* interaction types. In the envelope network, a link from i to j exists if it exists in at least one of the six relational networks.

Network Measures. The statistical properties of the six networks have been studied as separate entities using the following notations and measures. N_α is the number of nodes in the network type α , and $L_\alpha^{\text{dir(undir)}}$ is the number of (un)directed links. Reciprocity is labeled by r_α and $\rho(k_\alpha^{\text{in}}, k_\alpha^{\text{out}})$ is the correlation of in- and out-degrees within the α network. Average degree, clustering coefficient, and clustering coefficient with respect to the corresponding random graph are marked by \bar{k}_α , C_α , and $C_\alpha/C_\alpha^{\text{rand}}$, respectively. For more details, see the *SI Text*.

Network Interactions. For network–network interactions, we compute the Jaccard coefficient which measures the interaction between two networks by measuring the tendency that links simultaneously are present in both networks. $J_{\alpha\beta}$ is a similarity score between two sets of elements and is defined as the size of the intersection of the sets divided by the size of their union (50), $J_{\alpha\beta} \equiv |\alpha \cap \beta| / |\alpha \cup \beta|$. Related similarity measures, such as the cosine similarity measure lead to comparable results. The correlation measures used are described in detail in the *SI Text*.

The legal department of the Medical University of Vienna has attested the innocuousness of the used anonymized data.

ACKNOWLEDGMENTS. We thank Roberta Sinatra for helpful remarks. R.L. acknowledges support from the UK Engineering and Physical Sciences Research Council. This work was conducted within the framework of European Cooperation in Science and Technology Action MP0801 Physics of Competition and Conflicts. M.S. and S.T. acknowledge support from the Austrian Science Fund Fonds zur Förderung der wissenschaftlichen Forschung P 19132. During the redaction of this paper, we discovered an independent paper where the authors perform a large-scale verification of structural balance and arrive at conclusions similar to ours (51).

1. Wasserman S, Faust K (1994) Social network analysis: Methods and applications. (Cambridge Univ Press, Cambridge, UK), pp 37–248.
2. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. *Annu Rev Sociol* 27:415–444.
3. Entwisle B, Faust K, Rindfuss RR, Kaneda T (2007) Networks and contexts: Variation in the structure of social ties. *Am J Sociol* 112:1495–1533.
4. Padgett JF, Ansell CK (1993) Robust action and the rise of the Medici, 1400–1434. *Am J Sociol* 98:1259–1319 1400–1434.
5. Evans TS, Lambiotte R (2009) Line graphs, link partitions and overlapping communities. *Phys Rev E* 80:016105.
6. Ahn Y-Y, Bagrow JP, Lehmann S (2009) Link communities reveal multi-scale complexity in networks. *Nature*, in press.
7. Mucha PJ, Richardson T, Macon K, Porter MA, Onnela JP (2010) Community structure in time-dependent, multiscale, and multiplex networks. *Science* 328:876–878.
8. Selee TM, Kolda TG, Kegelmeyer WP, Griffin JD (2007) Exposing multi-relational networks to single-relational network analysis algorithms. *Comput Sci Res Inst Summer Proc* 2007:87–103.
9. Rodriguez MA, Shinauer J (2010) Extracting clusters from large datasets with multiple similarity measures using IMSCAND. *J Informetr* 4:29–41.
10. Baym NK, Zhang YB, Lin MC (2004) Social interactions across media. *New Media Soc* 6:299–318.
11. Huang Y, Shen C, Williams D, Contractor N (2009) Virtually there: Exploring proximity and homophily in a virtual world. *Proc 2009 Int Conf Comput Sci Eng* 4:354–359.
12. Genini J, Morelato LPC, Guimarães PR, Jr, Olesen JM (2010) Cheaters in mutualism networks. *Biol Lett*, in press.
13. Newman MEJ, Barabási A-L, Watts DJ (2006) *The Structure and Dynamics of Networks* (Princeton Univ Press, Princeton, NJ), pp 221–228.
14. Newman MEJ (2003) The structure and function of complex networks. *SIAM Rev* 45:167–256.
15. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D (2006) Complex networks: Structure and dynamics. *Phys Rep* 424:175–308.
16. Lazer D, et al. (2009) Computational social science. *Science* 323:721–723.
17. Onnela JP, et al. (2007) Structure and tie strengths in mobile communication networks. *Proc Natl Acad Sci USA* 104:7332–7336.
18. Szell M, Thurner S (2010) Measuring social dynamics in a massive multiplayer online game. *Soc Netw*, in press.
19. Newman MEJ, Forrest S, Balthrop J (2002) Email networks and the spread of computer viruses. *Phys Rev E* 66:035101.
20. Lambiotte R, et al. (2008) Geographical dispersal of mobile communication networks. *Physica A* 387:5317–5325.
21. Eagle N, Pentland A, Lazer D (2009) Inferring social network structure using mobile phone data. *Proc Natl Acad Sci USA* 106:15274–15278.

22. Castronova E (2005) *Synthetic Worlds: The Business and Culture of Online Games* (Univ of Chicago Press, Chicago), pp 1–284.
23. Bainbridge WS (2007) The scientific research potential of virtual worlds. *Science* 317:472–476.
24. Castronova E (2006) On the research value of large games. *Games Cult* 1:163–186.
25. Johnson NF, et al. (2009) Human group formation in online guilds and offline gangs driven by a common team dynamic. *Phys Rev E* 79:066117.
26. Easley D, Kleinberg J (2010) *Networks, Crowds, and Markets: Reasoning About a Highly Connected World* (Cambridge Univ Press, Cambridge, UK), in press.
27. Heider F (1946) Attitudes and cognitive organization. *J Psychol* 21:107–112.
28. Cartwright D, Harary F (1956) Structure balance: A generalization of Heider's theory. *Psychol Rev* 63:277–293.
29. Newman MEJ (2002) Assortative mixing in networks. *Phys Rev Lett* 89:208701.
30. Amaral LAN, Scala A, Barthelemy M, Stanley HE (2000) Classes of small-world networks. *Proc Natl Acad Sci USA* 97:11149–11152.
31. Garlaschelli D, Loffredo MI (2004) Patterns of link reciprocity in directed networks. *Phys Rev Lett* 93:268701.
32. Barabási A, Albert R (1999) Emergence of scaling in random networks. *Science* 286:509–512.
33. Coleman JS (1988) Social capital in the creation of human capital. *Am J Sociol* 94:S95–S120.
34. Rapoport A (1953) Spread of information through a population with socio-structural bias I: Assumption of transitivity. *B Math Biophys* 15:523–533.
35. Leicht EA, D'Souza RM (2009) Percolation on interacting networks. arXiv:0907.0894.
36. Haythornthwaite C (2001) Exploring multiplexity: Social network structures in a computer-supported distance learning class. *Inform Soc* 17:211–226.
37. Antal T, Krapivsky P, Redner S (2006) Social balance on networks: The dynamics of friendship and enmity. *Physica D* 224:130–136.
38. Marvel S, Kleinberg J, Strogatz S (2009) The energy landscape of social balance. *Phys Rev Lett* 103:198701.
39. Davis JA (1967) Clustering and structural balance in graphs. *Hum Relat* 20:181–187.
40. Christakis NA, Fowler JH (2007) The spread of obesity in a large social network over 32 years. *New Engl J Med* 357:370–379.
41. Bohorquez JC, Gourley S, Dixon AR, Spagat M, Johnson NF (2009) Common ecology quantifies human insurgency. *Nature* 462:911–914.
42. Lim M, Metzler R, Bar-Yam Y (2007) Global pattern formation and ethnic/cultural violence. *Science* 317:1540–1544.
43. Clauset A, Gleditsch KS (2009) The developmental dynamics of terrorist organizations. arXiv:0906.3287.
44. de Kerchove C, Van Dooren P (2008) The pagetrust algorithm: How to rank web pages when negative links are allowed. *Proc SIAM Int Conf Data Mining* 346–352.
45. Traag VA, Bruggeman J (2009) Community detection in networks with positive and negative links. *Phys Rev E* 80:036115.
46. Hamill JT, Deckro RF, Chrissis JW, Mills RF (2008) Analysis of layered social networks. *IO Sphere (Winter)* 27–33.
47. Carrington P, Scott J, Wasserman S (2005) *Models and Methods in Social Network Analysis* (Cambridge Univ Press, New York), pp 8–30.
48. Lewis K, Kaufman J, Gonzalez M, Wimmer A, Christakis N (2008) Tastes, ties, and time: A new social network dataset using facebook.com. *Soc Networks* 30:330–342.
49. Newman MEJ, Strogatz SH, Watts DJ (2001) Random graphs with arbitrary degree distributions and their applications. *Phys Rev E* 64:026118.
50. Traud AL, Kelsic ED, Mucha PJ, Porter MA (2008) Community structure in online collegiate social networks. arXiv:0809.0690.
51. Leskovec J, Huttenlocher D, Kleinberg J (2010) Signed networks in social media. *Proceedings of the 28th International Conference on Human Factors in Computing Systems* 1361–1370.

Supporting Information

Szell et al. 10.1073/pnas.1004008107

SI Text

This supplement to the paper “Multirelational organization of large-scale social networks in an online world” contains detailed information on the various network types analyzed, an overview of the game, notations of measures used, a detailed analysis of signed triad dynamics, and the null model and STC model for the social balance part.

Different Types of Social Interactions.

- *Friendship and enmity networks.* Players can anonymously mark others as friends or enemies, for any reason. The marked players are added to the marker's personal friends or enemies list. Additionally, every player has a personal “friend of” and “enemy of” list, displaying all players who have marked them as friend or enemy, respectively. Friend and enemy markings can be removed anytime.
- *Communication networks.* Private messages (PMs) are the prevalent form of communication within the game. It is similar to e-mail—a PM is only seen by sender and receiver.
- *Trade networks.* These are extracted by considering two kinds of trade possibilities between players: Either players meet and exchange game money and/or commodities, or players visit commercial outlets of other players and buy/sell commodities or equipment.
- *Attack networks.* For extracting attack networks, we select all attacks carried out by players on other players or on commercial outlets.
- *Bounty networks.* Links in bounty networks represent bounties, which are amounts of game money placed on other players or on their commercial outlets. Any player can collect a bounty by terminating the bountied player or destroying his commercial outlet.

How Players Get to Interact with Each Other. For a basic understanding of how the implementation of the game may shape the patterns of interactions between players, we give here an overview of the mechanisms and motivations leading players to get to know and to interact with each other. For more details see ref. 1.

Every game action carried out by a player (trade, movement, attack, etc.) costs a certain amount of so-called action points (APs). These points can not exceed a maximum value. For players having less APs than the maximum, every few minutes a small number of APs is replenished automatically. Once a player's character is out of APs, she has to wait to be able to play on. As a result, the typical Pardus player logs in once a day to spend all her APs on several activities within a few minutes. Social activities such as writing PMs do not consume APs.

A Pardus universe has the shape of a two-dimensional lattice (bounded in several ways) on which players can move (movement consumes action points). On each field (the smallest unit of this lattice), a player has the option to construct a building. Buildings act both as production sites and trade outlets for certain commodity types. Typically, a player has up to five buildings. Players may visit buildings of other players to trade game money for commodities or vice versa. A player has a trade tie with every other player who traded at her buildings, or whose buildings she traded with. Additionally, there is the (much rarer) possibility that two players meet on the same field and exchange game money and/or commodities. Attack comes with the same two options: Either a player attacks the building of another player or the player himself (for this interaction they again have to stand on the same field).

All other relations (communication, friendship, enmity, bounty) are independent of location in space, i.e., every player may write PMs to, mark as friend or enemy, or set a bounty on any other player at anytime, provided she knows the target's character name. This name is visible on the navigation screen when players stand on the same field, in an online list which shows all currently online players, in chat channels, and in the game's forums provided a player has posted a message in the corresponding place, as well as in several sections of the game such as on news pages. We suspect that the type of acquaintances a player makes during the course of the game depends strongly on her involvement in social activities. If the player does not show the preference of using provided ways of communication (PMs, chats, forums), her partners of friendship/enmity/bounty interaction are likely to show a high causal dependency with her visited locations in the game universe. On the other hand, a frequent use of communication tools may reduce this dependency, because then interactions take place with players independent of location.

Besides character names and online status being displayed on every player's personal PM contacts page for quick access, the friends and enemies lists serve game-mechanic purposes: Friends/enemies are automatically or optionally included/excluded for certain actions. For example, enemies of building owners are not able to use the services offered in the respective places. Note that friend and enemy markings need not necessarily denote *affective* friendships or enmity, they rather indicate a certain degree of cooperative or noncooperative stance motivated by affective and/or cognitive incentives. However, we assume these two motives to coincide to a considerable extent, e.g., it seems highly unlikely that someone marked as enemy/friend due to rational considerations at the same time constitutes the affective opposite of friend/enemy within the game (and vice versa).

We have no information about external forms of relation or communication, e.g., players being real-life friends or communicating via external tools. Further, so far we do not know how well structure and dynamics of different types of social networks in Pardus match comparable social networks in real life, with a few exceptions (1): We have shown good agreement of PM network features with properties of mobile phone call networks and revealed findings well according with classic sociological hypotheses. Further, we have begun studying positive and negative networks as single entities and found results highly consistent with social balance theory, as well as a coincidence of network properties (triad significance profiles) with nonvirtual social networks.

Comparison with Existing Datasets. To our knowledge, the only large-scale dataset incorporating multiple interactions is the Facebook network of Lewis et al. (2). Pardus offers several advantages over this Facebook data. The Facebook network consists of three types of interactions between users: declared friendship relationships, picture friendships (being tagged in an online photo by a user), and dorm roommate friendships. However, two of these three types of interactions lead to a tainted representation of the social system. First, the friendship network of Facebook is known to be biased by the visibility of the friends of a user on its webpage (6). In Pardus, friend and enemy lists are *completely private*, meaning that no one except the marking and marked players have information about positive or negative ties between them. Our data thus represents a more realistic social situation, in the sense that social ties are not immediately accessible to the public but need to be found out by communication with or by careful observation of others. Second, dorm membership is ob-

tained from the projection of a bipartite network. This procedure is known to distort the number of cliques in a network (3). For this reason, we only focus on one-to-one interactions between players and discard indirect interactions such as the participation to a chat. Finally, the Pardus dataset has the advantage to capture a broad range of social interactions, because the players are immersed in the game and therefore not only communicate with one another, but also engage in collaborative/antagonistic actions.

Mathematical notations—Overlap. The standard way to represent a network is through its adjacency matrix A_{ij} which, for an unweighted, undirected network, is a symmetric matrix whose elements A_{ij} are equal to one if there is a link between i and j , and zero otherwise. To incorporate the existence of multiple relations (multiplex networks), it is common to define the tensor $A_{ij;\alpha}$, sometimes called supersociomatrix (4). This tensor has dimension $N \times N \times R$, where N is the total number of nodes and R is the number of different link types between the same set of agents. For a fixed value of α , $A_{ij;\alpha}$ is the adjacency matrix of the network defined by link type α . By construction, the properties of each network can be obtained from its adjacency matrix $A_{ij;\alpha}$. For instance, the degree $k_{i;\alpha}$ of a node i is given by $\sum_j A_{ij;\alpha}$, the total number of links in network α is $L_\alpha = \sum_i k_{i;\alpha}/2$. The number of paths of length n between nodes i and j is given by $(A_\alpha^n)_{ij}$. A whole new layer of complexity opens once the interplay between different sorts of networks is considered. From a mathematical point of view, multiplexity can be revealed by coupling different adjacency matrices. For instance, the (link) overlap $O_{\alpha\beta} = \frac{1}{2} \sum_{ij} A_{ij;\alpha} A_{ij;\beta}$ between the graphs α and β counts the number of links they have in common. Similarly, the multiplicity $m_{ij} = \sum_\alpha A_{ij;\alpha}$ of a link between i and j counts the number of different links between these nodes.

Correlation measures. Several network measures are based on the Pearson's correlation between two quantities. For two random variables X and Y with mean values \bar{X} and \bar{Y} , and standard deviations σ_X and σ_Y , the correlation coefficient $\rho(X, Y)$ is defined as:

$$\rho(X, Y) = \frac{E[(X - \bar{X})(Y - \bar{Y})]}{\sigma_X \sigma_Y} \in [-1, 1]. \quad [\text{S1}]$$

The *reciprocity* coefficient r is the correlation coefficient between the transposed entries of the adjacency matrix of a directed graph, $X \equiv A_{ij}$, $Y \equiv A_{ji}$ (5). Similarly, we introduce the coefficients $\rho(k_\alpha^{\text{in}}, k_\alpha^{\text{out}})$ to evaluate the correlations between in-degree and out-degree around the same node in a graph α and $\rho(k_\alpha, k_\beta)$, to evaluate the correlations between the degrees of a node in the two different graphs α and β . The coefficient $\rho(k_\alpha^{\text{in}}, k_\alpha^{\text{out}})$ is a measure of the deviation of a directed network from a Eulerian network, i.e., $\rho(k_\alpha^{\text{in}}, k_\alpha^{\text{out}}) = 1$ only for a Eulerian graph, namely $k_{i;\alpha}^{\text{in}} = k_{i;\alpha}^{\text{out}}$ for each node i , while $\rho(k_\alpha, k_\beta)$ measures the correlation of the degree centrality of the same node in different networks. The coefficient $\rho(\text{rk}(k_\alpha), \text{rk}(k_\beta))$ is calculated the same way as $\rho(k_\alpha, k_\beta)$, with the difference that not degrees but ranks of degrees are used, i.e., the node with largest degree has rank 1, the second largest has rank 2, etc. Nodes with the same degree have the same rank; the difference to the subsequent rank is the number of nodes which shared the previous rank. For example if there are three nodes with degree 45 and rank 10, nodes with degree 44 have rank 13.

Relations between network-network measures. Correlating network-network measures reveals a strong relation between link overlap and degree correlation ($\rho = 0.88$, p -value: 10^{-5}), see SI Fig. 1 (a). Pairs of networks of the same connotation have a higher overlap than oppositely connotated pairs; a similar tendency for degree correlation is apparent. A correlation between

link overlap and degree rank correlation is also present, however with lower significance ($\rho = 0.63$, p -value: 0.01), see SI Fig. 1 (b). We mark pairs including a communication network as neutral, since messages may involve both positively or negatively connotated content.

Network-network interactions over time. To assess to what extent network-network properties of link overlap, degree correlation, and degree rank correlation change over time, we show these properties at days 150, 300, and 445 for all pairs of networks in SI Fig. 2. Here, accumulated networks, i.e., all except friendship and enmity networks, are accumulated over days 1 to 150, over days 1 to 300, and over days 1 to 445, respectively. Friendship and enmity networks are taken at these times. The number of players involved in the envelope network (i.e., in any relation) changes from 9,862 to 15,103 to 18,819 in these points in time, respectively. Changes are relatively small, except for degree correlations of pairs including bounty networks. Overlap values generally tend to decrease slightly over time.

Social balance and sparse networks. To analyze the multiplexity of large-scale networks and to draw conclusions from our observations, we need to address an issue that is usually obsolete for experiments on small social systems. When considering different types of interactions between students of a class or diplomatic positions between countries, it is reasonable to assume that all agents in the network are aware of each other's existence. In large-scale social networks, in contrast, the absence of any type of link between two nodes may either correspond to the existence of an indifferent/neutral interaction, or to the absence of any past and present contact between both agents. The fact that agents only know a fraction of the total set of agents is typical of sparse networks and originates from the finite capacities of its nodes, i.e., agents have limited time and resources, therefore can explore a small fraction of the available spatial and cognitive space. In the Pardus networks this finiteness is affirmed by the observation that out-degrees of friendship and enmity networks have an upper bound, limited by the Dunbar number of ≈ 150 (1), presumed as a natural limit for social ties humans are able to sustain (7). The average degree \bar{k}_α is well below $O(100)$, for all types α .

A proper null model. The aspect of a null model becomes important when assessing the relevance of topological structures in a network. A standard procedure consists in comparing this observation against similar observations in null models, i.e., randomized versions of the original network under adequate constraints (8). In order to test predictions of structural balance theory, we focus on friendship and enmity relations, and leave aside other types of interactions. In a first step, we remove the negligible number of ambiguous links (links between players where one marks the other as friend but is marked back as enemy). Our strategy is now to compare the numbers N_{Δ_i} of triads with i positive links with the expected numbers $N_{\Delta_i}^{\text{rand}}$ of triads in a null model. A standard choice for a null model consists of random graphs with fixed degree sequences. It has been applied for each network separately in Table 1, where we observe that friendship and enmity networks are both more transitive than a random graph. However, this choice is not appropriate to test the arrangement of positive and negative links on the set of existing relations between agents—a reshuffling of topology by keeping degrees fixed would for example considerably change the number of triads which we want to keep fixed. For this reason we define a null model by keeping the topology fixed and by randomly assigning the L_+ plus-signs and L_- minus-signs on the existing links, where L_+ (L_-) are the original numbers of friendship (enmity) links respectively. $N_{\Delta_i}^{\text{rand}}$ is measured by averaging over 1000 realizations of the null model. Moreover, the deviation of the data from randomness is evaluated by the so-called z-score:

$$z_i = \frac{N_{\Delta_i} - N_{\Delta_i}^{\text{rand}}}{\sigma_{\Delta_i}}, \quad [2]$$

where σ_{Δ_i} is the standard deviation of the number of triads Δ_i .

Given the ratio $p := \frac{L_+}{L_+ + L_-}$ of positive to all links in a signed network, the expected ratio of triad types in the sign-shuffled null model is, following straightforward combinatorial arguments, p^3 , $3(1-p)p^2$, $3(1-p)^2p$, and $(1-p)^3$ for triad types $+++$, $++-$, $+--$, $---$, respectively. These expressions were used to create Fig. 5 (center) in the main text.

Dynamics of signed triadic structures and network growth. By measuring all day-to-day transitions from wedges (triads with two links, with the possible forms $++$, $+-$, and $--$) to the other triadic structures ($+++$, $++-$, $+--$, $---$, $+++$, $+-$, $+$, $-$, and the empty triad) we shed light upon the mechanisms which lead to the observed significant social balance discussed in the main text. We measure the following possible transition types: A wedge stays the same, closes with a positive/negative link (with the original links unchanged), has one or both links removed, or has the sign of one or both links switched. SI Fig. 3 shows the daily transition probabilities, normalized by the total number of wedges of the corresponding type on that day. Due to lack of notability the transition type of switching links (with a probability less than a tenth of that of link-removal, on average) was not included in SI Fig. 3.

Wedges of type $++$ close preferentially with a positive link, see green line in SI Fig. 3 (a), wedges of type $+-$ with a negative link, see blue line in SI Fig. 3 (b). These probabilities are decreasing over time and seem to eventually level out. There is no clear sign preference in the closure of type $--$ wedges (red lines). These observations consistently explain the social balance results shown in Fig. 4 and Fig. 5 in the main text. Further, note that $--$ wedges are much more likely to remain unchanged than other types of wedges, see SI Fig. 3 (d). We conclude that the mechanism of triadic closure (9) has a much weaker influence as a driving force in purely negative tie networks than in positive tie or signed networks.

SI Fig. 4 depicts the total number of wedges of each type, for every day. Note how the majority of wedges is of type $--$, although there are more positive than negative links, see Table 1 (main text). Also the growth rate for $--$ wedges is higher than for the other two types (until about day 350, where the number of $--$ wedges starts to equilibrate). This seemingly paradoxical circumstance is consistent with the marked differences in clustering coefficients, see Table 1 (main text). It is further consistent with the observation that a number of aggressive players frequently offend many others and consequently get marked as enemy by unconnected players (1). Since the clustering coefficient measures the ‘closedness’ of triads, a high clustering coefficient in friendship networks implies a relatively small number of $++$ wedges, whereas a low clustering coefficient in enmity networks implies a relatively high number of $--$ wedges.

For assessing to what extent network growth is driven by the closure of triads, we define the *closure ratio* as the number of newly added links which close at least one wedge, divided by the number of all new links, over a certain time-window during the evolution of the network. The closure ratio lies between 0 and 1; the higher it is the more new links close a wedge. In practice, the closure ratio is strictly smaller than 1, since a number of cases unavoidably do not allow for the possibility of new links closing a wedge (for example the first and second links which are added into an empty network because no wedges exist at that stage). The measured time-evolution of daily closure ratios in the friend-enmity multiplex-network is depicted in SI Fig. 5 (a). Over time the ratio slightly increases and seems to level out at around 0.5 for both positive and negative links. We conclude that half of all links added close at least one wedge, while the other half does

not close one. Thus, a model for network growth using only wedge transition rates shown in SI Fig. 3 could only account for the dynamics of about half of the added links.

Another quantity important for modeling social network dynamics is the number of removed links per time. We define the *link churn* ch as the number of removed links divided by the number of new links, over a given time-window. The churn is nonnegative; there are 3 possible cases: *i*) Growth ($0 \leq ch < 1$): More new than removed links, *ii*) Equilibrium ($ch = 1$): The same number of new as removed links, *iii*) Shrinkage ($ch > 1$): More removed than new links. The higher ch , the more links are removed relative to the number of added links. Note that in the majority of classic network growth models, such as preferential attachment (10), no removal of links is assumed ($ch = 0$) and the effect of churn is ignored. The measured time-evolution of link churns over time windows of 14 days in the friend/enmity network is depicted in SI Fig. 5 (b). Over time ch increases and fluctuates around $ch = 0.7$ for both positive and negative links (taken over days 200 to 445, $ch = 0.66$ for friend links and $ch = 0.72$ for enemy links). Therefore, at the end for every three new links about two links are removed. Since the number of links removed from wedges is much smaller than links added to close wedges, we conclude that many links are removed from triadic structures other than wedges. Again, a model for network growth using only the transition rates shown in SI Fig. 3 would only account partially for link-removal dynamics.

A network evolution model of signed triadic closure. Using the measured daily transition rates above we define a transition matrix

$$M_{\text{STC}} = P \begin{pmatrix} ++ \rightarrow +++ & ++ \rightarrow ++- & ++ \rightarrow ++ \\ +- \rightarrow +-+ & +- \rightarrow +-- & +- \rightarrow +- \\ -- \rightarrow --+ & -- \rightarrow --- & -- \rightarrow -- \end{pmatrix} = \begin{pmatrix} 0.000212 & 0.000029 & 0.999759 \\ 0.000025 & 0.000279 & 0.999696 \\ 0.000040 & 0.000036 & 0.999924 \end{pmatrix},$$

where the entries are the probabilities of a wedge of given type to another triadic structure. Rows 1, 2, and 3 of M_{STC} distinguish between wedge types $++$, $+-$, and $--$, respectively; columns 1, 2 and 3 distinguish between probabilities for closure with either a positive or a negative link, or the probability of no change, respectively. The constant probabilities of columns 1 and 2 are determined by averaging the corresponding evolving probabilities over the days 100 to 445 (this time-window was chosen due to a relatively decreased level of fluctuations in transition probabilities, see SI Fig. 3). The third column is one minus the sum of values in column 1 and 2, since we neglect link-removals and sign-switches. With these parameters we design the following network evolution model, to understand *Signed Triadic Closure* (STC):

- At time t pick wedge i at random (random sequential update)
- Determine the type of wedge i and close (or do not close) it according to the relevant entry in M_{STC}
- Pick next wedge until all wedges are updated
- Continue with time step $t + 1$

As initial condition we take the observed friendship and enmity multiplex-network at day 100. Simulating this process leads to the results shown in Fig. 5 (right) in the main text, reproducing the ratio of triads in the data considerably better than the null model.

For the purpose of simplicity, the STC model ignores three possibly important aspects:

- It does not take into account links added by means other than triadic closure. As we have shown above, the closure ratio ≈ 0.5 , i.e., only half of all new links are added in the process of triadic

closure. For the other half of new links, one could for example take into account the mechanism of preferential attachment (10), for which exponents were reported in (1).

- The model does not take into account the removal of links. Because removal happens in a relatively high frequency ($\text{ch} \approx 0.7$), a possible extension of the model could involve the measurement and implementation of decay rates, i.e., of

transition rates from complete signed triads to wedges (or other triadic structures with a smaller number of links). We suspect that balanced triads are more stable than unbalanced triads.

- The model could be expanded to incorporate directed links and in/out-degree distributions, clustering coefficients, assortativity, etc.

1. Szell M, Thurner S (2009) Measuring social dynamics in a massive multiplayer online game. arXiv:0911.1084.
2. Lewis K, Kaufman J, Gonzalez M, Wimmer A, Christakis N (2008) Tastes, ties, and time: a new social network dataset using facebook.com. *Social Networks* 30:330–342.
3. Newman MEJ, Strogatz SH, Watts DJ (2001) Random graphs with arbitrary degree distributions and their applications. *Phys Rev E* 64:026118.
4. Wasserman S, Faust K (1994) Social network analysis: methods and applications (Cambridge University Press).
5. Garlaschelli D, Loffredo MI (2004) Patterns of link reciprocity in directed networks. *Phys Rev Lett* 93:268701.
6. Golder S, Wilkinson D, Huberman B (2007) Rhythms of social interaction: Messaging within a massive online network. In *Communities and Technologies 2007: Proceedings of the Third Communities and Technologies Conference* (Springer, New York).
7. Dunbar R (1993) Co-evolution of neocortex size, group size and language in humans. *Behavioral and Brain Sciences* 16:681–735.
8. Maslov S, Sneppen K (2002) Specificity and stability in topology of protein networks. *Science* 296:910–913.
9. Rapoport A (1953) Spread of information through a population with socio-structural bias I: assumption of transitivity. *Bull Math Biophys* 15:523–533.
10. Barabási A, Albert R (1999) Emergence of scaling in random networks. *Science* 286:5439–509.

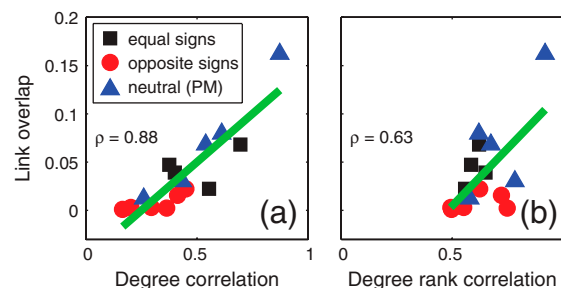


Fig. S1. Correlation between the network-network measures of (a) link overlap and degree correlation (p -value: 10^{-5}), (b) link overlap and degree rank correlation (p -value: 0.01). Pairs of networks with equal connotation (positive-positive or negative-negative) have a higher overlap than oppositely signed pairs, with a similar tendency for degree correlation. We mark pairs including a communication network as neutral, since messages may involve both positively or negatively connotated content.

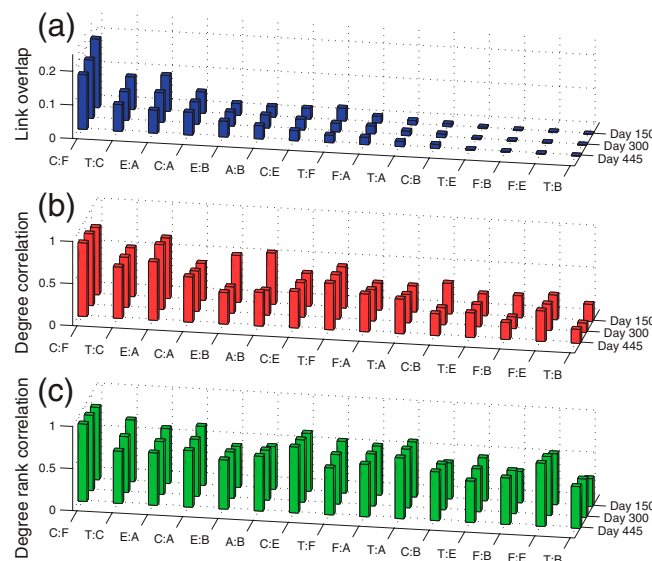


Fig. S2. Evolving network-network properties (a) link overlap, (b) degree correlation, (c) degree rank correlation, on the days 150, 300, and 445, for all pairs of networks, with the notations E for Enmity, F for Friendship, A for Attack, T for Trade, C for Communication and B for Bounty. Changes are relatively small, except for degree correlations in the E:B, A:B, C:B, and F:B interactions.

