How women organize their social networks different from men

Supplementary Information

Michael Szell, Stefan Thurner

CONTENTS

This document is the Supplementary Information (SI) for the manuscript *How women organize their social networks different from men.* Besides the definitions of measures for network related quantities it contains a series of information on the complete multiplex data.

S1. DEFINITION OF NETWORK MEASURES

In the following we give detailed definitions and properties for all measures used in the article and the SI.

Graph

In mathematical terms, networks are described by *graphs* [3, 4]. An undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$ is defined as a pair of sets, the node set $\mathcal{N}$ containing all nodes n_i and the link set $\mathcal{L}$ containing unordered pairs $l_{ij} := \{n_i, n_j\}$ denoting those nodes which are connected by an undirected link (edge). A directed graph (*digraph*) has a link set $\mathcal{L}$ which contains *ordered* pairs $l_{ij} := (n_i, n_j)$ marking nodes which are connected by a directed link (arc) going from n_i to n_j. The expression N denotes the cardinality of the set $\mathcal{N}$.

Symmetrization

The *symmetrization* of a digraph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$ is constructed as follows: Start with $\mathcal{G}^* = (\mathcal{N}, \mathcal{L}^*)$, where $\mathcal{L}^*$ is an empty link set, and for all pairs of nodes n_i and n_j add the undirected link l_{ij} to $\mathcal{L}^*$ if the directed link $l_{ij} \in \mathcal{L}$ or if $l_{ji} \in \mathcal{L}$. By L^{dir} and L^{undir} we denote the cardinalities of the sets $\mathcal{L}$ and $\mathcal{L}^*$, respectively.

Weighted network

In unweighted graphs all links are treated equally. A *weighted graph* is a generalization in which the weight w_{ij} of a link l_{ij} may take any non-zero real value.

Degree

In an undirected graph the *degree* k_i of a node n_i is the number of links connecting to it. All k_i nodes which are directly linked to n_i are called (nearest) *neighbors* of n_i. We denote the average degree of all nodes in a network by $\bar{k}$. In a directed graph the in-degree k_i^{in} of a node n_i is the number of its incoming links, the out-degree k_i^{out} the number of its outgoing links.

Neighbor degree

We denote the average degree of all nearest neighbors of a node n_i by $\bar{k}_i^{\mathrm{nn}}$. We denote the average degree of all nearest neighbors of all nodes as a function of degree k by $\bar{k}^{\mathrm{nn}}(k)$.

Correlation between in and out degrees

We write $\rho = \rho(k^{\mathrm{in}}, k^{\mathrm{out}})$ for the correlation of in- and out degrees within the α network.

Clustering coefficient

The *clustering coefficient* C_i of node n_i in an undirected graph is the ratio between the number y_i of links between its k_i neighbors and the number of all possible links $k_i(k_i-1)/2$ between them,

$$C_i := \frac{2y_i}{k_i(k_i-1)}. \tag{1}$$

The network's clustering coefficient C is the average over all clustering coefficients, $C = (1/N)\sum_i C_i$. A random graph's clustering coefficient C_r is given by $C_\mathrm{r} = \bar{k}/N$ [4].

Reciprocity

Reciprocity measures the tendency of individuals to reciprocate connections, i.e. the creation of mutual instead of asymmetric dyads [3]. A naive reciprocity index can be defined by

$$R := \frac{L^{\mathrm{dir}}}{L^{\mathrm{undir}}} - 1, \tag{2}$$

where L^{undir} is the number of undirected links in the symmetrization of the digraph. Values of $R = 0$ and $R = 1$ stand for no mutual dyads and mutual dyads only, respectively.

Triad

A *triad* is a (sub)graph consisting of three nodes. In a digraph there exist 16 isomorphism classes of triads [5]. We adopt the notation of [6] for the 13 connected classes, i.e. for the classes having no isolated nodes.

Triad significance profile and Z-score

The *triad significance profile* (TSP) is the vector of statistical significances of each triad class compared to random networks drawn from the $U(X_{*+}, X_{+*}, M^*)$ distribution, i.e. of random networks having identical in/out degrees and equally likely numbers of mutual dyads for each node [6, 7]. Statistical significance of a triad class i is measured by the Z score

$$Z_i = \frac{(N_i^{\mathrm{real}} - \bar{N}_i^{\mathrm{rand}})}{\mathrm{std}(N_i^{\mathrm{rand}})}, \tag{3}$$

where N_i^{real} is the frequency of occurence of the triad class in the considered network, and $\bar{N}_i^{\mathrm{rand}}$ and $\mathrm{std}(N_i^{\mathrm{rand}})$ are the average frequency of occurence and the standard deviation in an ensemble of random networks drawn from $U(X_{*+}, X_{+*}, M^*)$. The TSP is the normalized vector of all 13 Z scores,

$$\mathrm{TSP}_i = \frac{Z_i}{\left(\sum_{i=1}^{13} Z_i^2\right)^{1/2}} \tag{4}$$

Jensen-Shannon divergence

We use the Jensen-Shannon divergence

$$S_{rel}(p,q) = \frac{1}{2}\left[\sum_{k=1}^{max} p(k)\log\frac{2p(k)}{p(k)+q(k)} + \sum_{k=1}^{max} q(k)\log\frac{2q(k)}{p(k)+q(k)}\right] \tag{5}$$

to compare degree distributions, where p and q are the distributions of male, female or male control players, either for in or out degree distributions. The Jensen-Shannon divergence is a standard symmetric measure for comparing probability distributions and is based on the Kullback-Leibler entropy [8].

S2. NETWORK PROPERTIES

Supplementary Table I shows a series of properties of the single network layers. The majority of properties shows no substantial gender-specific difference, except for reciprocity R (friends and trades), average degree $\bar{k}$ (PMs and trades), average nearest neighbor degree $\bar{k}^{nn}$ (PMs, enemies), clustering coefficient C (trades, attacks). Concerning significance in terms of rejecting the null hypothesis H_0 of equal means (see Methods in the main text), we find significant rejection similarly for average degree $\bar{k}$ (both 2 sigmas for PMs and trades), average nearest neighbor degree $\bar{k}^{nn}$ (PMs, 3 sigmas), clustering coefficient C (trades, 5 sigmas). The t-test cannot be applied to the aggregated reciprocity values R.

S3. TIME-TO-RESPOND FOR ALL ACTIONS

We report the cumulative distributions of times-to-respond for all six action types in Supplementary Fig. 1. Most FM distributions are identical to their MF counterparts, except for friends and enemies (see main text).

S4. TRIAD CENSUS AND TRIAD SIGNIFICANCES

Supplementary Table II shows the census of undirected triads with genders on day 856 and corresponding Z-scores. Due to different numbers of males and females, values between genders are not directly comparable. Also, the Z-scores of triads scale in a non-trivial way with network size. However, some qualitative differences in Z-score signs – where significant – can give clues on different networking effects between genders. We make the following observations: Due to the small number of females, there exists even a much smaller number of FFF triads compared to MMM triads. Values of MMM and FFF Z-scores are qualitatively similar, except for (i) a significant underrepresentation of $---$ FFF triads (-3.13) as opposed to $---$ MMM triads (however, because there are only 9 $---$ FFF triads this may be due to statistical fluctuation), (ii) the Z-score of the asymmetric $+--$ FMM triad (31.43) is roughly twice the Z-score of the symmetric $+--$ FMM triad (15.69), reflecting the combinatorial fact that twice as many asymmetric triads are expected, but standing in contrast to the Z-scores of asymmetric and symmetric $+--$ MFF Z-scores which are almost the same (7.27 and 6.99, respectively).

S5. RELATIVE-DEGREE DISTRIBUTIONS

When directly comparing degree distributions between males and females, differences or similarities are not feasible to assess, see the cumulative in- and out-degree distributions in Supplementary Fig. 2. Because of this we consider the (pointwise) relative differences between the cumulative male and female degree distributions. Supplementary Fig. 3 shows for each degree k the relative difference between the value $P(\geq k)$ in the male and the female distribution, for both in- and out-degrees and all six types of networks.

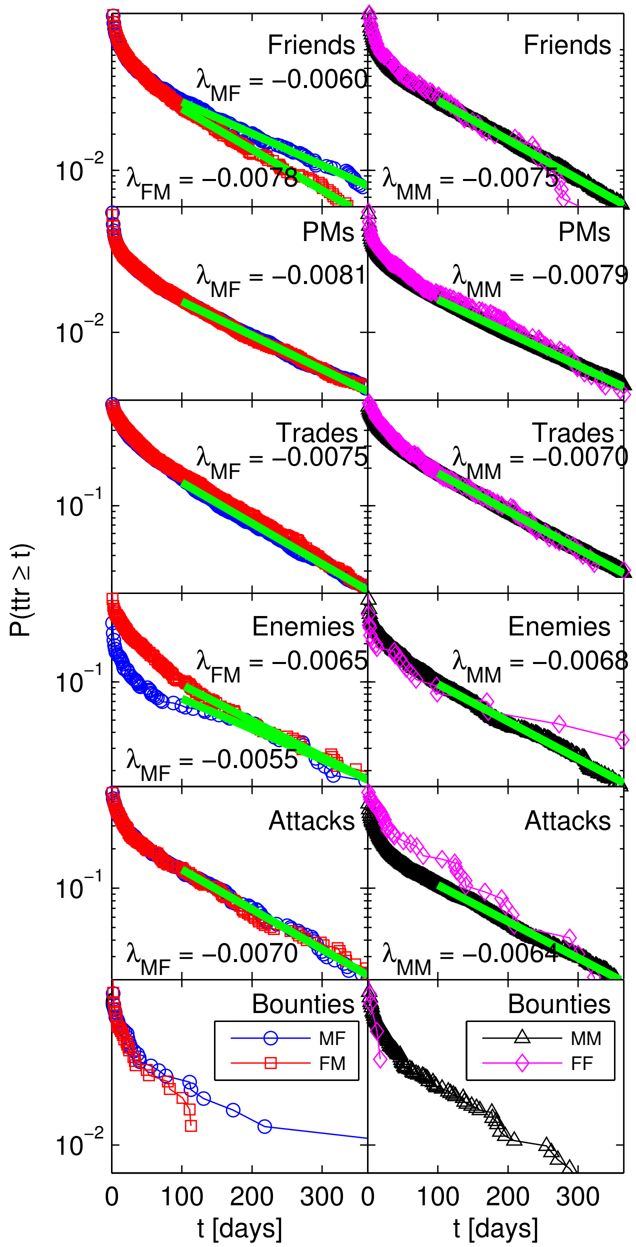
S6. GENDER DIFFERENCES IN NW-NW INTERACTIONS

To quantify the resulting inter-dependencies between pairs of networks, we follow the two approaches as in Ref. [2]. 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. The Jaccard coefficient quantifies 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, $J_{\alpha\beta} \equiv |\alpha \cap \beta|/|\alpha \cup \beta|$. As a second measure, we compute correlations $\rho(k_\alpha, k_\beta)$ between node degrees in different networks. 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 1, 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 $\rho(\mathrm{rk}(k_\alpha), \mathrm{rk}(k_\beta))$ between rankings of node degrees. Overlap and correlation quantities provide complementary insights into the organization of social structures. In Supplementary Figs. 4, 5 and 6 for all pairs of networks the three measures are shown, either for all players, or males or females only, respectively. Orderings here are slightly different from the orderings in Ref. [2] because the trade relation (T) used here is based

Supplementary Table I. Properties of the six networks on day 856, split into all possible classes of genders of male (M), female (F) or all, or into gender-gender interactions of male-male (MM), male-female (MF), female-male (FM), female-female (FF) where applicable. Upper part, directed networks: Number of nodes N, number of directed links L^{dir}, reciprocity R, correlation of in and out degrees ρ, Jensen-Shannon divergence of degree distributions S_{rel}. Lower part, symmetrized undirected networks: Number of undirected links L^{undir}, degree $\bar{k}$, nearest neighbor degree $\bar{k}^{nn}$, clustering coefficient C. Substantial differences are: Higher reciprocity R in FF friend and trade links compared to MM links, higher average degree $\bar{k}$ of females in PM and trade networks compared to males, lower average nearest neighbor degree $\bar{k}^{nn}$ of females in PM and enemy networks compared to males, higher clustering coefficient C of females in trade and attack networks compared to males. For values with stars the null hypothesis H_0 of equal means is rejected. Deviations from the male values denote standard deviations using 8 random male control groups each having the same number of male players as the female group, see main text.

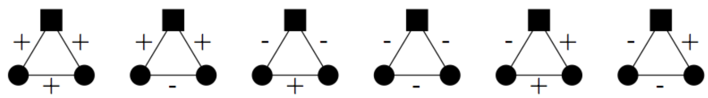
	Positive ties			Negative ties		
Directed	Friends	PMs	Trades	Enemies	Attacks	Bounties
N_α All	4469	22540	14825	3047	11381	4362
N_α M	3838	20045	13055	2657	10078	3819
N_α F	631	2495	1770	390	1303	543
L_α^{dir} All	42357	551293	138882	24219	102467	9284
L_α^{dir} MM	31449	422732	104477	18146	81654	7066
L_α^{dir} MF	4968	60378	15916	2984	10648	994
L_α^{dir} FM	5048	58646	15722	2651	8892	1103
L_α^{dir} FF	892	9537	2767	438	1273	121
R_α All	0.60	0.81	0.31	0.13	0.15	0.21
R_α MF	0.63	0.81	0.32	0.12	0.15	0.24
R_α MM	0.59 $\pm$0.03	0.81 $\pm$0.01	0.31 $\pm$0.02	0.13 $\pm$0.07	0.13 $\pm$0.02	0.20 $\pm$0.04
R_α FF	0.68	0.83	0.38	0.08	0.15	0.20
ρ_α All	0.85	0.97	0.84	0.14	0.72	0.42
ρ_α M	0.79	0.97	0.83	0.02	0.71	0.39
ρ_α F	0.88	0.98	0.82	-0.00	0.75	0.35
S_{rel}^{in} MF	0.0234	0.0296	0.0144	0.0651	0.0142	0.0099
S_{rel}^{out} MF	0.0272	0.0367	0.0297	0.0287	0.0373	0.0183
Undirected						
$L_\alpha^{\mathrm{undir}}$ All	29580	327838	117055	22675	94767	8300
$L_\alpha^{\mathrm{undir}}$ MM	22117	251634	88270	16944	75516	6347
$L_\alpha^{\mathrm{undir}}$ MF	6875	70607	26542	5311	18074	1844
$L_\alpha^{\mathrm{undir}}$ FF	588	5597	2243	420	1177	109
$\bar{k}_\alpha$ All	13.23	29.09	15.79	14.88	16.65	3.81
$\bar{k}_\alpha$ M	13.32 $\pm$0.67	28.63 $\pm$1.93	15.56 $\pm$0.72	14.75 $\pm$1.99	16.78 $\pm$1.58	3.81 $\pm$0.69
$\bar{k}_\alpha$ F	12.76	32.79 *	17.53 *	15.77	15.68	3.80
$\bar{k}_\alpha^{nn}$ M	41.54 $\pm$0.98	378.57 $\pm$3.99	81.93 $\pm$0.94	91.75 $\pm$2.62	107.91 $\pm$1.87	76.31 $\pm$6.12
$\bar{k}_\alpha^{nn}$ F	40.97	359.95 **	80.71	85.63	109.52	84.10
C_α All	0.231	0.256	0.141	0.034	0.069	0.020
C_α M	0.229 $\pm$0.009	0.255 $\pm$0.006	0.137 $\pm$0.006	0.0339 $\pm$0.003	0.0686 $\pm$0.004	0.0202 $\pm$0.004
C_α F	0.245	0.261	0.170 ****	0.0351	0.0761	0.022

* p-value $<$ 0.05, ** p-value $<$ 0.01, *** p-value $<$ 0.001 **** p-value $<$ 0.0001

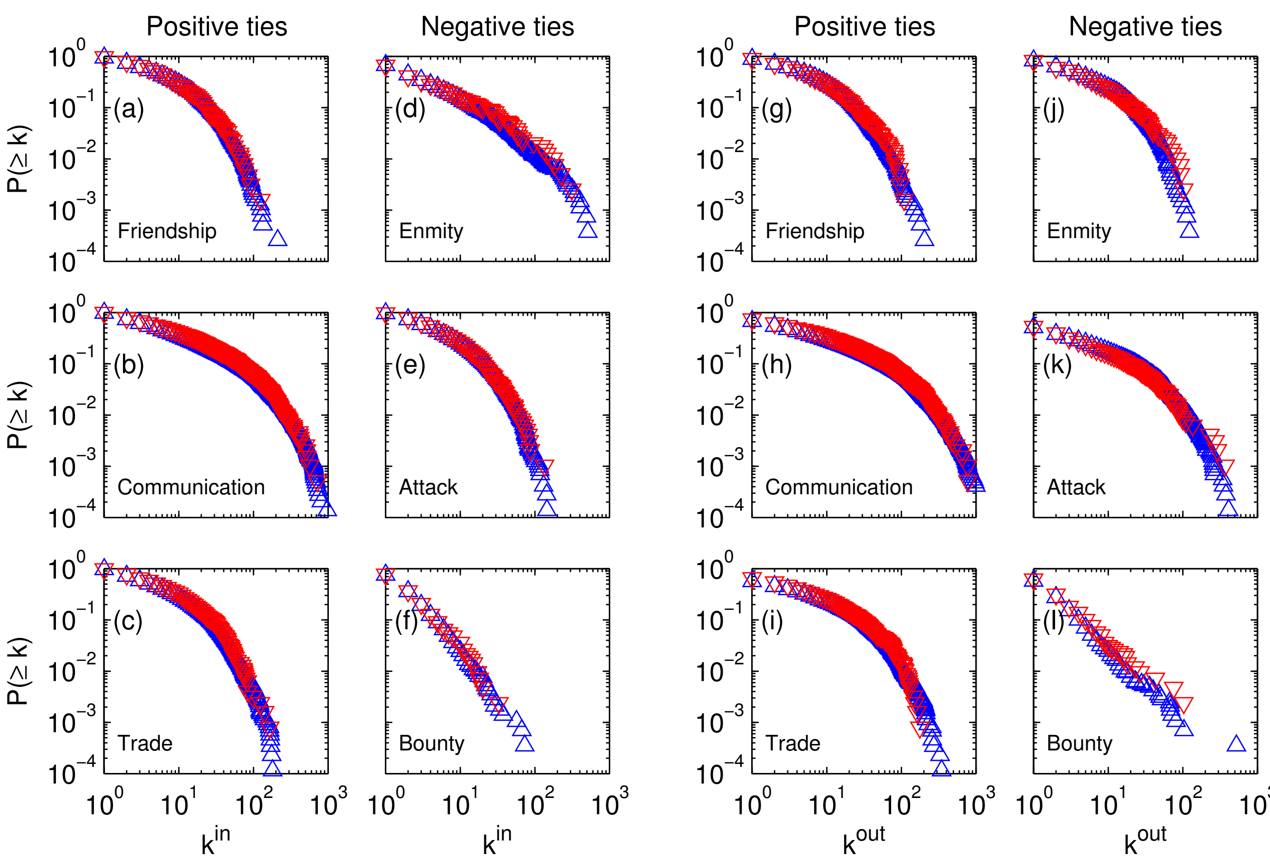


Supplementary Figure 1. (Left column) Cumulative probability distributions for the time-to-respond for females to reciprocate for all link types, given the initiator was male (MF), and vice versa (FM). (Right column) Situation for equal sex reciprocation MM, and FF. Shown are all six different relation types with exponential fits where feasible (green curves, fit ranges from day 100 to 365).

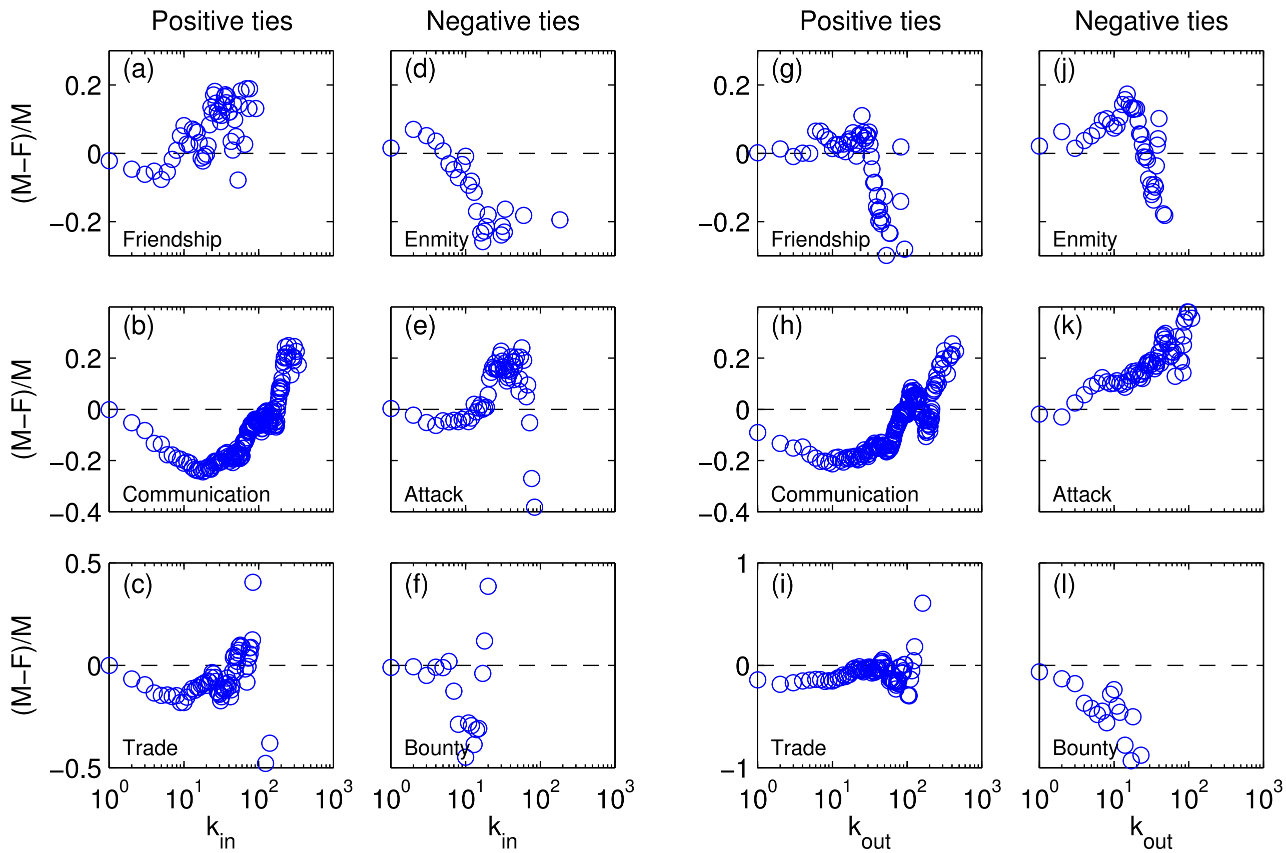
Supplementary Table II. Triad census of signed (friend and enmity) undirected triads with genders, on day 856, and corresponding Z-scores using 1000 random reshuffles of signs. There are 4646 nodes in this multiplex network, 3992 males and 654 females. For rows 3 and 4, the columns 2 and 3 show symmetric cases where the single player with the different gender has two links of the same kind. For rows 3 and 4, the columns 5 and 6 show the corresponding asymmetric cases where the single player with the different gender has one positive and one negative link.



	Census					
MMM	31076	5854	37498	6672	—	—
FFF	181	27	185	9	—	—
FMM	14134	1004	6512	3089	2150	11476
MFF	2586	186	1014	382	335	2045
	Z-score					
MMM	56.53	-136.96	39.80	0.46	—	—
FFF	9.34	-12.87	4.46	-3.13	—	—
FMM	38.90	-26.89	15.69	-0.24	-70.17	31.43
MFF	20.05	-11.35	6.99	-3.54	-16.38	7.27



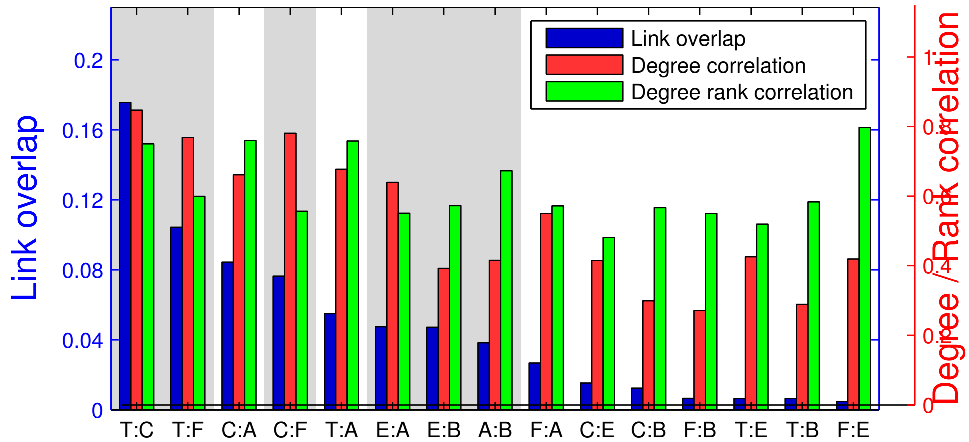
Supplementary Figure 2. (a)–(f) Cumulative in-degrees of males and fermales, (g)–(l) cumulative out-degrees of males and females.



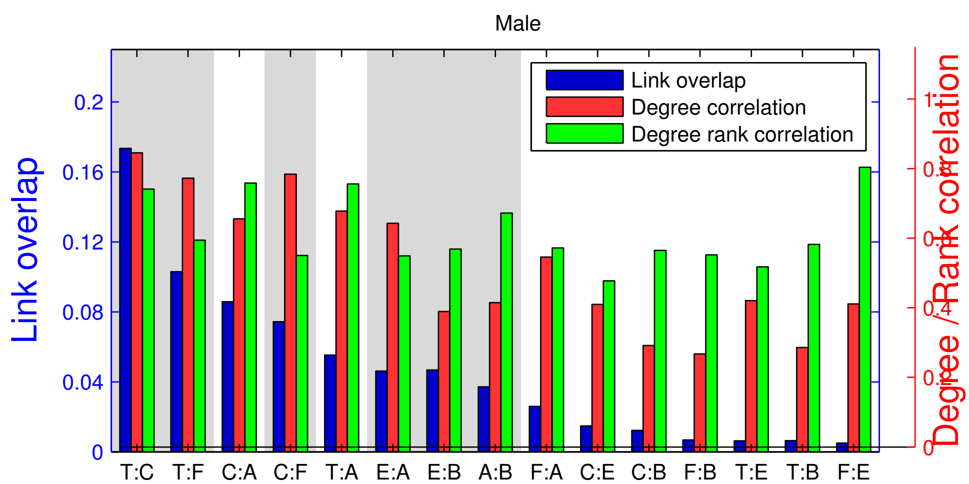
Supplementary Figure 3. (a)–(f) Relative differences between in-degree distributions, (g)–(l) relative differences between out-degree distributions, i.e. for each degree k (in and out respectively) we calculate the relative difference (M-F)/M between the male and female degree distributions.

only on ship-to-ship trades as opposed to all trades in [2]. Supplementary Fig. 7 shows the relative differences (m-f)/m between male and female network-network measures.

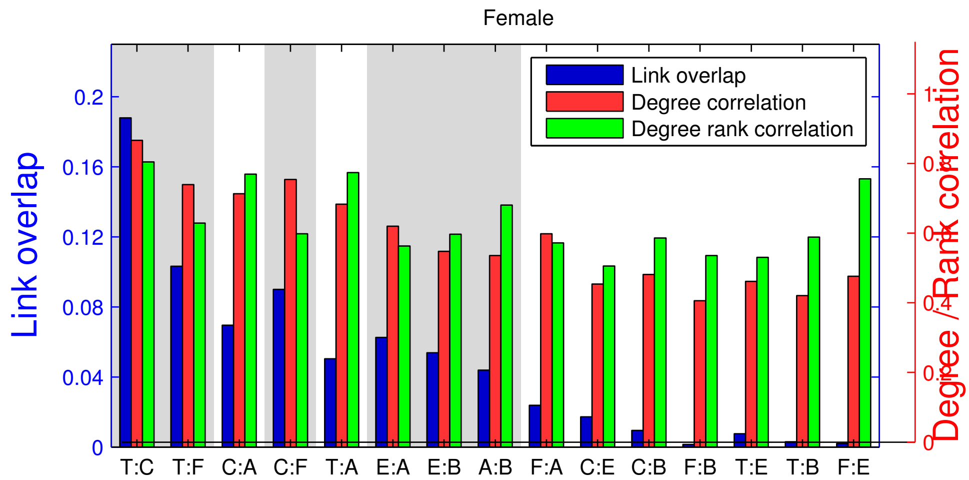
[1] Szell, M. & Thurner, S. Measuring social dynamics in a massive multiplayer online game. *Social Networks* **32**, 313–329 (2010).
[2] Szell, M., Lambiotte, R. & Thurner, S. Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences* **107**, 13636–13641 (2010).
[3] Wasserman, S. & Faust, K. *Social Network Analysis: Methods and Applications* (Cambridge University Press, 1994).
[4] Dorogovtsev, S. & Mendes, J. *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, 2003).
[5] Harary, F., Norman, R. & Cartwright, D. *Structural Models: An Introduction to the Theory of Directed Graphs* (John Wiley & Sons Inc, 1965).
[6] Milo, R. *et al.* Network motifs: Simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
[7] Roberts, J. Simple methods for simulating sociomatrices with given marginal totals. *Social Networks* **22**, 273–283 (2000).
[8] Cha, S. Comprehensive survey on distance/similarity measures between probability density functions. *International journal of mathematical models and methods in applied sciences* **1**, 300–307 (2007).



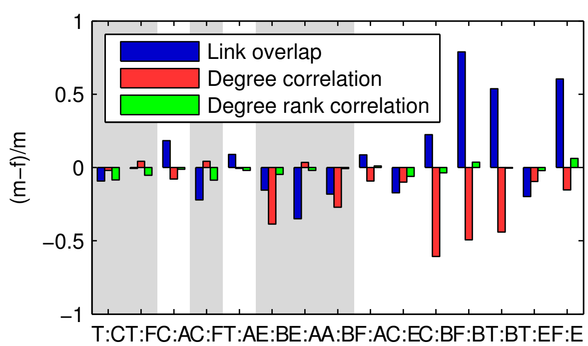
Supplementary Figure 4. Link overlap (Jaccard coefficient), degree correlation $\rho(k_\alpha, k_\beta)$ and degree rank correlation $\rho(\mathrm{rk}(k_\alpha), \mathrm{rk}(k_\beta))$ for all pairs of networks (ordered by link overlap), with the notations E for Enmity, F for Friendship, A for Attack, T for Trade, C for Communication and B for Bounty. Pairs of equal connotation (positive-positive or negative-negative) are marked with a gray background. These pairs have high overlaps, while oppositely-connotated pairs have lower overlaps.



Supplementary Figure 5. Same plot as in Supplementary Fig. 4 (using the same ordering), but for male players only.



Supplementary Figure 6. Same plot as in Supplementary Fig. 4 (using the same ordering), but for female players only.



Supplementary Figure 7. Relative differences between male and female network-network measures.