

SCIENTIFIC REPORTS



OPEN

Skill complementarity enhances heterophily in collaboration networks

Received: 07 October 2015

Accepted: 25 November 2015

Published: 08 January 2016

Wen-Jie Xie^{1,2,3}, Ming-Xia Li^{2,3}, Zhi-Qiang Jiang^{1,4}, Qun-Zhao Tan⁵, Boris Podobnik^{6,7,8,9,10}, Wei-Xing Zhou^{1,3,4} & H. Eugene Stanley⁶

Much empirical evidence shows that individuals usually exhibit significant homophily in social networks. We demonstrate, however, skill complementarity enhances heterophily in the formation of collaboration networks, where people prefer to forge social ties with people who have professions different from their own. We construct a model to quantify the heterophily by assuming that individuals choose collaborators to maximize utility. Using a huge database of online societies, we find evidence of heterophily in collaboration networks. The results of model calibration confirm the presence of heterophily. Both empirical analysis and model calibration show that the heterophilous feature is persistent along the evolution of online societies. Furthermore, the degree of skill complementarity is positively correlated with their production output. Our work sheds new light on the scientific research utility of virtual worlds for studying human behaviors in complex socioeconomic systems.

Complexity emerges in the evolving and self-organizing processes of many natural, social, technological, and biological systems. The constituents of a complex system interact with each other and form complex evolving networks, where the constituents are nodes and their interaction relationships are links^{1–6}. For many real networks, the link formation process follows either the global principle of popularity in which a node tends to link with high-degree nodes^{7,8}, or the local principle of similarity in which a node tends to link with nodes having traits similar to its own⁹, or a tradeoff between them⁹.

In the sociological literature the local principle of similarity, i.e., the phenomenon that “birds of a feather flock together,” is known as homophily¹⁰. There is much empirical evidence indicating that individuals prefer to forge social ties with people whose traits such as education, race, age, and sex are the same as their own^{11–14}. Such homophilous behaviors are ubiquitous in social networks and have been well documented^{10–12,14–18}. In addition, the similarity shared by individuals in a group is often a significant predictor of a group’s altruism level and its ability to cooperate¹⁹. Sociological literature argues that human societies tend to display two social systems: (i) homophilous, in which people seek out people who are similar, and (ii) heterophilous, in which people seek out people who are different²⁰. The evidence indicating the actual existence of heterophilous societies is rare, however. One example is to study team formation processes in offline gangs and online games depending on the heterogeneity of agents’ attributes²¹.

In general, it has long been accepted that one of the most significant factors in increasing productivity in modern human societies has been the division of labor²². Thus we might assume that people in modern societies now prefer to forge links or collaborate with those who have complementary productive skills and that socioeconomic networks are becoming increasingly heterophilous, but no direct evidence of this has been documented. The availability of big data recorded from massively multiplayer online role-playing games (MMORPGs) enables us

¹School of Business, East China University of Science and Technology, Shanghai 200237, China. ²Postdoctoral Research Station, East China University of Science and Technology, Shanghai 200237, China. ³Department of Mathematics, East China University of Science and Technology, Shanghai 200237, China. ⁴Research Center for Econophysics, East China University of Science and Technology, Shanghai 200237, China. ⁵Shanda Games Ltd., 690 Bibo Road, Shanghai 201203, China. ⁶Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215, USA. ⁷Zagreb School of Economics and Management, 10000 Zagreb, Croatia. ⁸Luxembourg School of Business, Luxembourg. ⁹Faculty of Civil Engineering, University of Rijeka, 51000 Rijeka, Croatia. ¹⁰Faculty of Economics, University of Ljubljana, 1000 Ljubljana, Slovenia. Correspondence and requests for materials should be addressed to W.-X.Z. (email: wxzhou@ecust.edu.cn) or H.E.S. (email: hes@bu.edu)

to test social and economic hypotheses and theories—such as this one—in large-scale virtual populations²³ and gain a deeper understanding of our social and economic behaviors^{24–32}.

In this work, we study the collaboration formation process of individuals with different professional skills. A mathematical model is proposed by assuming that individuals in socioeconomic systems choose collaborators that are of maximum utility. Based on the evolving collaboration networks of 124 virtual worlds in which the agents (virtual people) belong to three different professions possessing different skills, empirical analysis and model calibration unveil that the agents prefer to collaborate with others of different professions. We further construct two measures to quantify the degree of complementarity of virtual societies. We find that social complementarity positively correlates with economic output.

Results

A model of collaboration formation. Consider a society or a community s on day t , whose size N_s is the number of s -agents. The number of (s, i) -agents is denoted by $N_{s,i}$, where $i = 1, 2$, and 3 stand for the three professions. Hence $N_s = \sum_{i=1}^3 N_{s,i}$. The ratio of i -agents in society s is

$$w_{s,i} = N_{s,i}/N_s. \quad (1)$$

The average number of j -collaborators of an (s, i) -agent is $f_{s,ij}$. Hence, the average number of collaborators that an (s, i) -agent has is $f_{s,i} = \sum_{j=1}^3 f_{s,ij}$. The average proportion of j -collaborators in all collaborators of an i -agent is

$$q_{s,ij} = \frac{f_{s,ij}}{\sum_{j=1}^3 f_{s,ij}} = \frac{f_{s,ij}}{f_{s,i}}. \quad (2)$$

Note that $q_{s,ij}$ is the homophily index^{15,33}. If i -agents have zero preference for collaborating with j -agents, we have $q_{s,ij} = w_{s,j}$. If i -agents prefer to collaborate with j -agents, we have $q_{s,ij} > w_{s,j}$. In this case the i -agents are homophilous when $j = i$ and the i -agents are heterophilous when $j \neq i$.

An agent seeks collaborators when she/he finds it difficult to complete a task alone. If there is no collaboration preference, the proportion of (s, j) -collaborators that an (s, i) -agent has is identical to the proportion of j -agents in the group, that is $q_{s,ij} = w_{s,j}$. Hence the number of (s, j) -collaborators of an (s, i) -agent is $f_{s,ij}^{\text{rnd}} = f_{s,i} w_{s,j}$. However, in a society with a division of labor, the choice of collaborators has a significant influence on the completion of the task and it is better to have collaborators with complementary skills. Therefore, the number and skill configuration (or distribution) of an agent's collaborators are the main determinants of her utility. We assume that, for an (s, i) -agent, there is an optimal configuration of collaborators with different skills, $f_{s,ij}^{\text{opt}} = \gamma_{ij} f_{s,i} w_{s,j}$, where the preference coefficients γ_{ij} are independent of society s . If the skill configuration in the collaborator list of an agent is optimal, her utility reaches its maximum. If the skill configuration deviates from that optimal value, her utility is reduced. In other words, the utility of an (s, i) -agent increases when her/his real number $f_{s,ij}$ of (s, j) -collaborators approaches the optimal value $f_{s,ij}^{\text{opt}}$ and reaches its maximum $U_{s,i}^{\max} = b f_{s,i}^\beta$ when her/his collaborator configuration is optimal such that $f_{s,ij} = f_{s,ij}^{\text{opt}}$. According to the law of diminishing marginal utility, we have $\beta < 1$. Therefore, the utility function of an (s, i) -agent is

$$U_{s,i} = b f_{s,i}^\beta - a f_{s,i}^\alpha X_{s,i}^{\beta-\alpha}, \quad (3)$$

where

$$X_{s,i} = \left[\sum_{j=1}^3 (q_{s,ij} - \gamma_{ij} w_{s,j})^2 \right]^{\frac{\alpha}{2(\beta-\alpha)}}, \quad (4)$$

in which γ_{ij} is the preference of (s, i) -agents for (s, j) -agents and $\alpha > 0$ since the second term in Eq. (3) quantifies the amount of utility decrease that is proportional to the deviation of the real configuration to the optimal configuration. If i -agents do not have any preference on j -agents such that $q_{s,ij} = w_{s,j}$ for all societies, we have $\gamma_{ij} = 1$. If i -agents prefer j -agents, we have $\gamma_{ij} > 1$. If i -agents prefer not to collaborate with j -agents, we have $\gamma_{ij} < 1$. For $\{i, j, k\} = \{1, 2, 3\}$, if $\gamma_{ij} > \gamma_{ik}$, then i -agents prefer j -agents over k -agents. To maintain a collaboration network of size $f_{s,i}$ the (s, i) -agent suffers a cost proportional to $f_{s,i}^{12}$,

$$M_{s,i} = c f_{s,i}^\beta. \quad (5)$$

According to the above model, the overall utility in the decision-making process is

$$D_{s,i}(f_{s,i}) = U_{s,i} - M_{s,i} = (b_s - c_s) f_{s,i}^\beta - a_s f_{s,i}^\alpha X_{s,i}^{\beta-\alpha}. \quad (6)$$

By maximizing $D_{s,i}(f_{s,i})$, we can estimate the parameters γ_{ij} (see *Materials and Methods*).

Empirical analysis. Figure 1A shows the collaboration networks on day $t = 15$ of a group of 27 agents randomly chosen from a virtual society filtered by three intimacy thresholds $I_c = 0, 100$, and 2000 . There are 12 warriors, 5 priests, and 10 mages. If i -agents are homophilous (neutral, heterophilous) in their collaboration-forging process, the proportion of links between i -agents is greater than (equal to, less than) the square of the proportion of i -agents (0.1975 for warriors, 0.0343 for priests, and 0.1372 for mages). For $I_c = 0$, there are 77 links

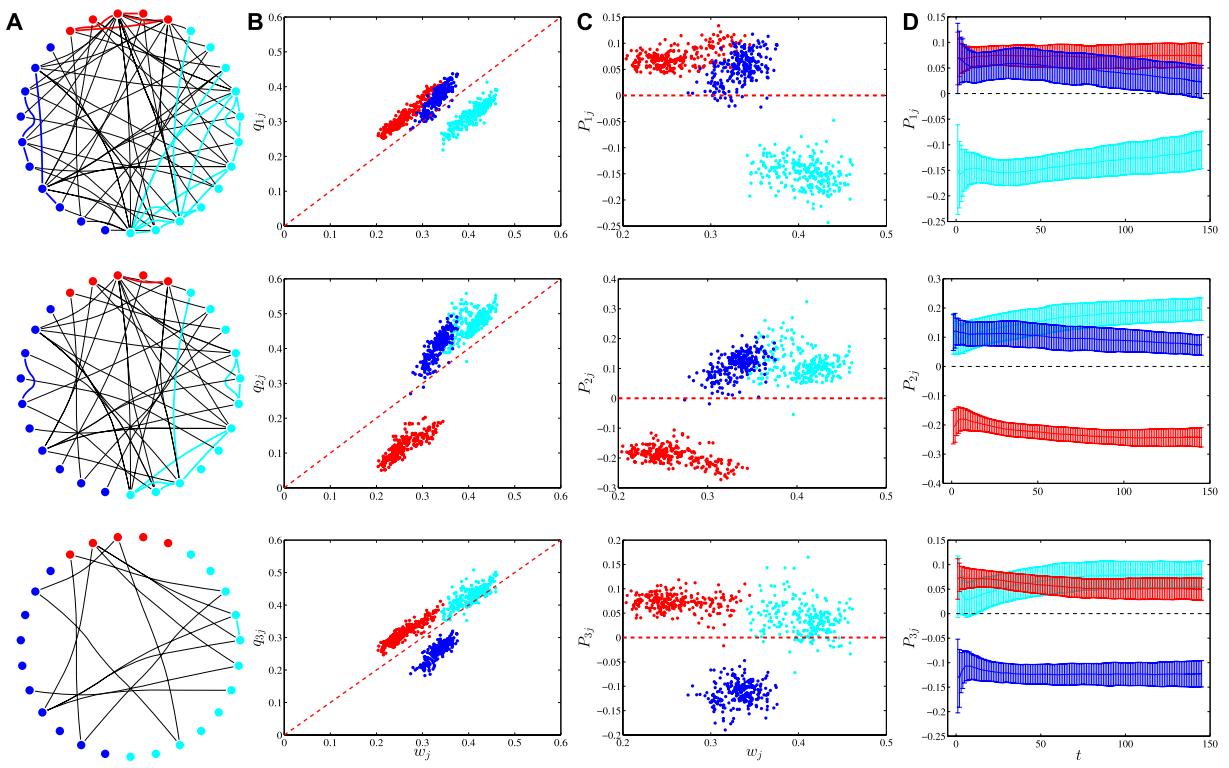


Figure 1. Empirical evidence of heterophily in the socioeconomic networks of virtual societies on a typical day $t = 15$. Warriors, priests and mages are marked respectively in cyan, red and blue. (A) Networks of 27 agents randomly chosen from a virtual society filtered by three intimacy thresholds $I_c = 0, 100$ and 2000 (top to bottom). (B) Dependence of $q_{s,ij}$ on relative size w_{sj} for all virtual societies for $I_c = 100$. In each plot, there are three well isolated clusters. For most societies, $q_{s,ij} > w_{sj}$ when $i \neq j$ and $q_{s,ij} < w_{sj}$ when $i = j$. (C) Dependence of preference measure $P_{s,ij}$ on relative size w_{sj} for all societies for $I_c = 100$. There are also three well separated clusters in each plot. For most societies, $P_{s,ij} > 0$ when $i \neq j$ and $P_{s,ij} < 0$ when $i = j$. (D) Evolution of the averaged preference measure $P_{s,ij}$ over all virtual societies for $I_c = 100$. The preference measures are roughly persistent.

including 15 intra-warrior links, 4 intra-priest links, and 4 intra-mage links. The proportions of intra-profession links are 0.1948 for warriors, 0.0519 for priests, and 0.0519 for mages. For $I_c = 100$, there are 48 links including 8 intra-warrior links, 1 intra-priest link, and 1 intra-mage link. The proportions of intra-profession links are 0.1667 for warriors, 0.0208 for priests, and 0.0208 for mages. For $I_c = 2000$, there are 15 links including only one intra-warrior link and no intra-priest and intra-mage links. The proportions of intra-profession links are 0.0667 for warriors and 0 for priests and mages. Hence, the agents in Fig. 1A are heterophilous except for priests when $I_c = 0$. We will show below that heterophily is not a specific characteristic for these 27 agents but a universal feature presents in all the virtual societies.

Figure 1B shows that when $t = 15$ and $I_c = 100$ most virtual societies have $q_{s,ij} > w_{sj}$ when $i \neq j$, but $q_{s,ij} < w_{sj}$ when $i = j$. Such heterophilous patterns are observed for other values of t and I_c as well (see Fig. S1).

Similar to the inbreeding homophily index^{15,33}, we define the collaboration preference index to be

$$P_{s,ij} = \frac{q_{s,ij} - w_{sj}}{1 - w_{sj}}. \quad (7)$$

Note that $P_{s,ii}$ is the inbreeding homophily value^{15,33}. If i -agents have no preference to collaborate with j -agents, we have $P_{s,ij} = 0$. If i -agents prefer to collaborate with j -agents, we have $P_{s,ij} > 0$. In the latter case, the i -agents are homophilous when $j = i$ and heterophilous when $j \neq i$. Empirical results show that for most virtual societies $P_{s,ij} > 0$ when $i \neq j$, but $P_{s,ij} < 0$ when $i = j$ (Fig. 1C and Fig. S2). Thus in socioeconomic networks the agents are heterophilous.

Figure 1D shows the evolution of preference values $P_{s,ij}$ averaged over all societies on the same day for $I_c = 100$. Although these curves exhibit mild trends, it is evident that the heterophilous feature is persistent as the virtual societies develop (see also Fig. S3).

Quantifying collaboration preference. To calibrate the model, we follow and further develop the econometric method presented in ref. 12 (see *Materials and Methods*). We obtain the values of γ_{ij} for each intimacy threshold I_c on each day t . Figure 2A shows the evolution of preference coefficients γ_{ij} for socioeconomic networks using the intimacy threshold $I_c = 100$, and Fig. 2B shows the average preference coefficients over all days. More

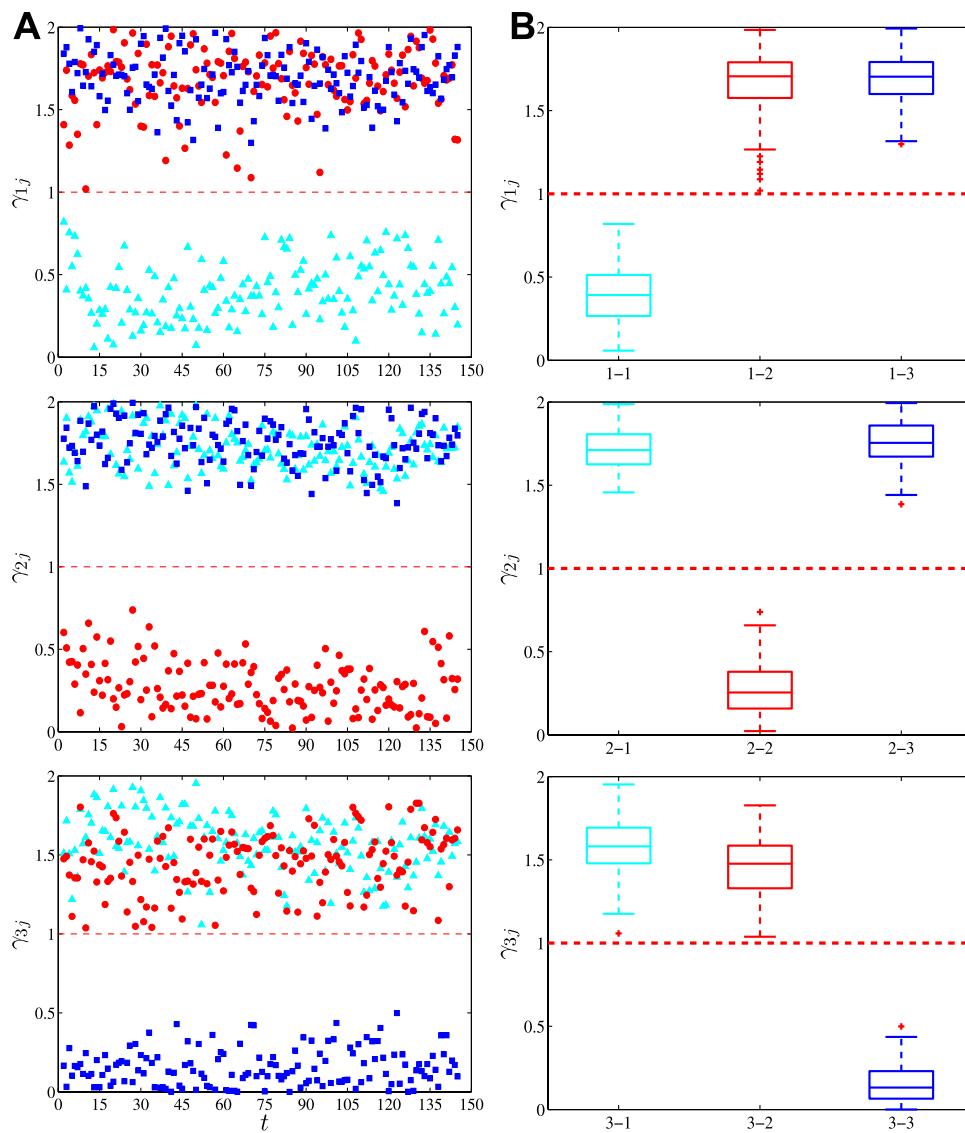


Figure 2. Preference coefficients γ_{ij} for socioeconomic networks with the intimacy threshold being $I_c = 100$. (A) Daily evolution of the nine preference coefficients γ_{ij} with $i, j \in \{1, 2, 3\}$. The color of a point (t, γ_{ij}) is determined by j : cyan, red and blue for $j = 1, 2$ and 3, respectively. The nine points for a given t were determined simultaneously in one calibration. (B) Box plots of γ_{ij} shown in (A).

results are given in Fig. S4 and Fig. S5 for $I_c = 0, 1, 10, 500, 1000$, and 2000 . The F -tests presented in *Materials and Methods* show that all the results are significant at the 0.1% level (see *SI Tables*).

All the estimated values of the γ_{ii} coefficients are less than 1, while all the γ_{ij} values for $i \neq j$ are greater than 1. This indicates that the agents are not seeking same-profession agents but different-profession agents and are thus heterophilous. In most cases, especially when the intimacy threshold I_c is not large, the $\gamma_{ij}(I_c, t)$ values do not have a trend along the evolution of virtual worlds. When I_c is large, however, we observe an increasing trend in $\gamma_{13}(I_c, t)$ for $I_c = 1000$ and 2000 , in $\gamma_{23}(I_c, t)$ for $I_c = 1000$ and 2000 , and in $\gamma_{32}(I_c, t)$ for $I_c = 500, 1000$ and 2000 (Fig. S4). We find that the preference coefficients might change with the increase of I_c (Fig. S4 and Fig. S5). For warriors, γ_{11} and γ_{13} decreases, while γ_{12} increases. For priests, γ_{21} increases, γ_{22} does not exhibit evident trend, while γ_{23} decreases. For warriors, γ_{31} increases, γ_{32} decreases, while γ_{12} increases for large I_c values.

There are also intriguing patterns of relative collaboration preference as quantified by $\gamma_{ij} - \gamma_{ik}$ where i, j and k correspond to the three professions (Fig. 2B and Fig. S5). On average, warriors prefer priests over mages and this relative preference enhances when I_c becomes greater but reduces slightly when i increases for large I_c values. Priests prefer mages over warriors when I_c values are small and prefer warriors over mages when I_c values are large. For large I_c , priests' relative preference on warriors over mages decreases along time t . Mages prefer priests over warriors when I_c is small and prefers warriors over priests when I_c is large. For large I_c , mages' relative preference on warriors over priests also decreases along time t .

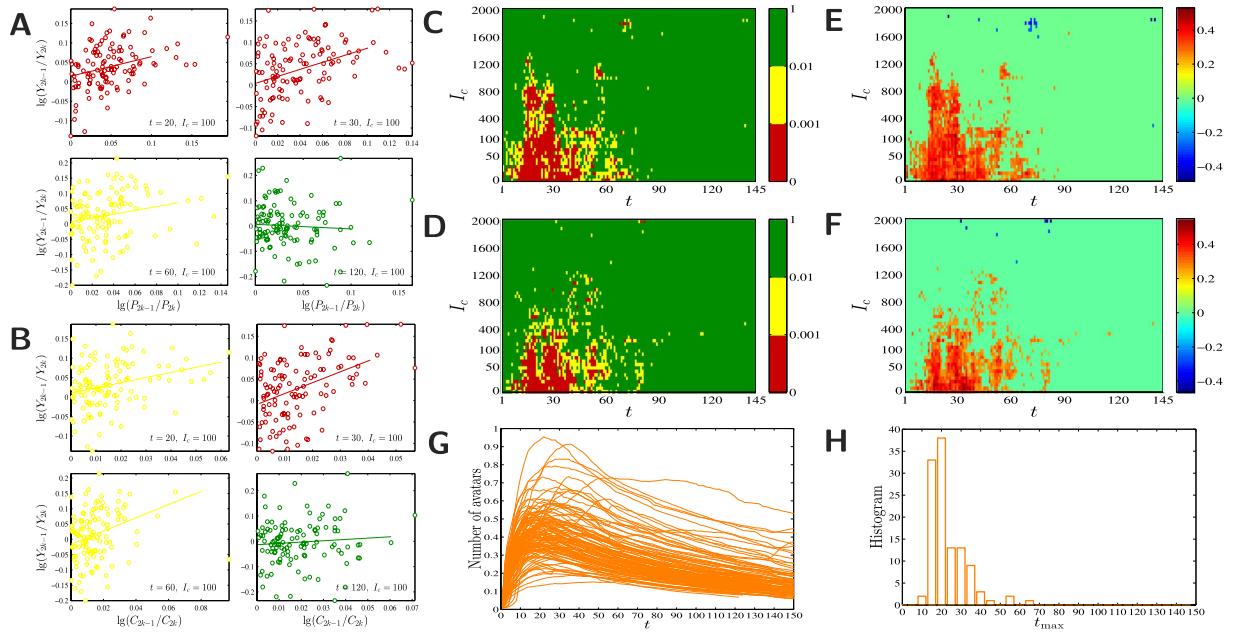


Figure 3. Relation between complementarity of collaboration network and economic output. (A) Examples of correlations between $\lg(P_{2k-1}/P_{2k})$ and $\lg(Y_{2k-1}/Y_{2k})$. (B) Examples of correlations between $\lg(C_{2k-1}/C_{2k})$ and $\lg(Y_{2k-1}/Y_{2k})$. (C) The p -value of the correlation between $\lg(P_{2k-1}/P_{2k})$ and $\lg(Y_{2k-1}/Y_{2k})$ for different values of I_c and t (in units of days). A give grid (t, I_c) is colored as red or yellow if the correlation is significant at the 0.001 level or the 0.01 level. Otherwise, the grid is colored as green. (D) The p -value of the correlation between $\lg(C_{2k-1}/C_{2k})$ and $\lg(Y_{2k-1}/Y_{2k})$. (E) Correlation coefficient ρ between $\lg(P_{2k-1}/P_{2k})$ and $\lg(Y_{2k-1}/Y_{2k})$ for different values of I_c and t . The correlation coefficient is set to be zero is the correlation is insignificant at the 0.01 level. (F) Correlation coefficient ρ between $\lg(C_{2k-1}/C_{2k})$ and $\lg(Y_{2k-1}/Y_{2k})$. (G) Evolution of the number of active agents in different virtual worlds. (H) Histogram of t_{\max} which is the date that a virtual world has historically the maximum active agents.

Group complementarity and economic output. To measure the economic implications of heterophilous preference in socioeconomic networks, we investigate the relationship between complementarity of professions and economic performance. Consider the socioeconomic network $\mathcal{N}_s(I_c, t)$ of a virtual society with intimacy threshold I_c on day t . Economic production utilizes virtual money and goods that are converted to a standardized currency (see *Materials and Methods*). For each member agent a in $\mathcal{N}_s(I_c, t)$, we calculate her production output in the week from $t - 6$ to t , denoted as $Y_{s,a}(t)$. The economic performance of the agents in $\mathcal{N}_s(I_c, t)$ is defined as the output per capita,

$$Y_s(I_c, t) = \frac{1}{\#\mathcal{N}_s(I_c, t)} \sum_{a \in \mathcal{N}_s(I_c, t)} Y_{s,a}. \quad (8)$$

One measure of profession complementarity can be defined as the sum of preference measures between the three types of agents,

$$P_s(I_c, t) = \sum_{j \neq i} P_{s,ij}(I_c, t). \quad (9)$$

Alternatively, we can measure complementarity by determining how much the real collaborator configuration $q_{s,ij}$ deviates from the optimal collaborator configuration $\gamma_{ij}w_{s,j}$ (see *Materials and Methods*). The lower the deviation, the higher the degree of complementarity. Thus, we have

$$C_s(I_c, t) \equiv \left[\sum_{i=1}^3 \sum_{j=1}^3 (q_{s,ij} - \gamma_{ij}w_{s,j})^2 \right]^{\frac{1}{2}}. \quad (10)$$

To make these results comparable for different virtual worlds, we investigate the relative quantities between two societies in the same world, $\lg(P_{2k-1}/P_{2k})$, $\lg(C_{2k-1}/C_{2k})$ and $\lg(Y_{2k-1}/Y_{2k})$, rather than focusing on each society separately. Both measures of complementarity correlate strongly with the relative economic output when t and I are not large (Fig. 3A–F, Fig. S7, and Fig. S8). For the first few days (small t), most agents strive to achieve higher levels by implementing specific tasks with small economic outputs. Other agents attempt to obtain high intimacy levels by killing monsters in locations unrelated to economic outputs. In both cases the agents intend to form complementary collaboration networks, but their activities are not focused on economic outputs. With the

development of a virtual world, the number of active agents increases and reaches a maximum at time t_{\max} and then decays (Fig. 3G). When the activity level of a virtual world decreases, the intent of the agents moves away from production and the collaboration structure is increasingly unrelated to economic activities. This is consistent with the fact that the spectrum of t_{\max} has a distribution similar to the significant correlations between complementarity and economic output (Fig. 3H).

Discussion

Overwhelming empirical evidence has shown that most social networks are homophilous. The probability that two nodes will connect is higher if they share similar traits. Our analysis of virtual worlds in which division of labor is operative demonstrates the important role of complementarity. In those socioeconomic networks individuals have the motivation to cooperate, and in the formation of the network individuals exhibit a heterophilous preference for those with complementary productive skills. Although mapping human behavior in virtual worlds to real-world human behavior is a subtle process³⁴, we believe that they share an intrinsic commonality because agents in virtual worlds are, in fact, controlled by real-world people. In particular, agents consciously form teams to accomplish tasks more successfully and effectively. More generally, growing evidence shows significant similarities in the behaviors of online agents and real-world humans^{23,35–43}.

In reality, human's preference is multidimensional in their traits¹³. The situation in virtual societies is a little different. Indeed, the way people interact with each other has significantly changed from the old days, particularly due to the impact of the Internet. In the modern time, people can meet through the Internet in the virtual world instead of physically getting together to dine, drink, and talk to forge ties. Personal traits become less important in virtual societies while agents' profession skill is identified as a dominating trait in virtual societies. Like most MMORPGs, the system is set up in a way that a party requires different roles to function optimally. In this sense, the main result of the paper would primarily reflect the design decisions of the game developers. On the other hand, however, such a setup is trying to mimic the real human society, in which people have different and diverse skills and hence there appears the division of labor²². Hence, the results documented in this work have a general significance.

The economic model proposed in this work is different from the one in ref. 12. The essential difference is in the assumption of the utility function. The choice of the utility function may have significant impact on the outcome of the model. We calibrated the original model in ref. 12 and the estimates of parameters suggested a homophilous behavior, which is inconsistent with the empirical results presented in Fig. 1. Also, we have used a modified method of model calibration. Moreover, our model allows us to determine not only if i -agents are homophilous or heterophilous but also the preference of one type of agents to any other type of agents. Hence, our model is more general and can be applied to other systems.

The relationship between social networks and economic output has been studied previously. It has been found, for example, that the diversity of individual relationships within a community strongly correlates with the economic development of the community⁴⁴ and is directly associated with higher productivity for both individuals and the community^{45,46}. Because, to date, detailed real data at the population level of societies have been unavailable, this correspondence between professional skill and economic performance has not been quantified. Here we have begun to fill this data gap and also to highlight the usefulness of virtual worlds in carrying out research in economics and sociology²³. One potential implication of our findings is that if a team leader or a firm manager recruits new members according the complementarity of their skills, the team's productivity will increase and the firm's economic well-being grow.

Materials and Methods

Data description. We use a huge database recorded from $K = 124$ servers of a popular MMORPG in China to uncover the patterns characterizing virtual socioeconomic networks. In a virtual world residing in a server there are two opposing camps or societies. Two agents can choose collaborators, and a measure of closeness called intimacy is assigned to the collaboration link. When two collaborators in the same society collaborate to accomplish a task, their intimacy level increases. Two agents from different societies can also collaborate, but their intimacy level remains zero. Hence the social networks of the two camps are essentially separate. We can regard the two camps as two societies, thus giving us $S = 248$ virtual societies. For convenience, $s = 2k - 1$ and $s = 2k$ stand respectively for the two societies in the same virtual world k . Two agents are defined as collaborators if they both are on the collaborator list and their intimacy exceeds I_c . We consider many temporal collaboration networks. On day t in a virtual society s , a network $\mathcal{N}_s(I_c, t)$ is a network in which the intimacies of all edges are no less than a threshold I_c , which can be disconnected (Fig. 1A).

There are a lot of different types of tasks in the virtual societies, which are embedded for agents of all levels. In some levels, the system will ask the agents to kill given amounts of different types of monsters. In other levels, the agents are asked to deliver something to a specific NPC (not-a-person character). And so on so forth. These tasks are usually not easy for the associated agents. However, they can ask their collaborators for help to form a team and fulfill the tasks together. Agents can also form teams to kill monsters and make productions. All these collaborations will increase the intimacy of the collaborating agents in the same team.

In each society there are three professions (warrior, priest, and mage). We use subscripts 1, 2, and 3 to stand respectively for the three professions: warrior, priest and mage. For simplicity, we define several notations as follows. An **s -agent** is an agent belonging to society s . An **i -agent** is an agent having profession i . Similarly, an **i -collaborator** is a collaborator having profession i . An **(s, i) -agent** or **(s, i) -collaborator** is an i -agent or i -collaborator in society s .

Model calibration. An (s, i) -agent solves the following decision-making problem of how many collaborators to have

$$\max_{f_{s,i}} D_{s,i}(f_{s,i}). \quad (11)$$

It follows that

$$f_{s,i} = [a_s \alpha / (b_s - c_s) \beta]^{\frac{1}{\beta-\alpha}} X_{s,i}. \quad (12)$$

Note that the γ_{ij} values are affected only by the professions and remain the same for different societies. This enables us to estimate the parameters.

The solution (12) denotes the average behavior (decision) of all agents having the same profession in a given society. If we consider an arbitrary agent a , we must add a noise term¹²,

$$f_{s,i;a} = [a_s \alpha / (b_s - c_s) \beta]^{\frac{1}{\beta-\alpha}} X_{s,i} + \varepsilon_a, \quad (13)$$

which means that the “realized” number of collaborators agent a has is the sum of a universal (or systemic) term and an idiosyncratic error term. The error term is assumed to have mean 0 and variance σ^2 . Note that this assumption states that the variance of any agent of any profession is the same.

We denote N_s as the size of society s and $w_{s,i}$ as the fraction of i -agents in society s . Hence the number of i -agents in society s is $N_s w_{s,i}$, and the expectation of the aggregated number of collaborators that i -agents have in society s is $N_s w_{s,i} f_{s,i}$. According to Eq. (13), we have

$$N_s w_{s,i} f_{s,i} = N_s w_{s,i} \left[\frac{a_s \alpha}{(b_s - c_s) \beta} \right]^{\frac{1}{\beta-\alpha}} X_{s,i} + E_{s,i}, \quad (14)$$

where $E_{s,i}$ has mean 0 and variance $(N_s w_{s,i} \sigma)^2 = (N_s w_{s,i})^2 \sigma^2$. It follows that, for $i \neq j$,

$$\frac{w_{s,i} N_s f_{s,i} - E_{s,i}}{w_{s,i} N_s X_{s,i}} = \frac{w_{s,j} N_s f_{s,j} - E_{s,j}}{w_{s,j} N_s X_{s,j}}. \quad (15)$$

Following ref. 12, we obtain an error for society s :

$$\begin{aligned} \Psi_{s,ij} &= w_{s,i} N_s f_{s,i} w_{s,j} N_s X_{s,j} - w_{s,j} N_s f_{s,j} w_{s,i} N_s X_{s,i} \\ &= N_s^2 w_{s,i} w_{s,j} (f_{s,i} X_{s,j} - f_{s,j} X_{s,i}) \end{aligned} \quad (16)$$

$$= E_{s,i} w_{s,j} N_s X_{s,j} - E_{s,j} w_{s,i} N_s X_{s,i}. \quad (17)$$

According to Eq. (17), we find that the mean of $\Psi_{s,ij}$ is 0 and the variance is $\phi_{s,ij} \sigma^2$, where

$$\begin{aligned} \phi_{s,ij} &= (w_{s,i} N_s)^2 (w_{s,j} N_s X_{s,j})^2 + (w_{s,j} N_s)^2 (w_{s,i} N_s X_{s,i})^2 \\ &= N_s^4 w_{s,i}^2 w_{s,j}^2 [(X_{s,i})^2 + (X_{s,j})^2] \end{aligned} \quad (18)$$

Thus the normalized variable $\Psi_{s,ij}^2 / \phi_{s,ij}$ has mean 0 and variance σ^2 for any society s . The sum of squared errors (Q_{ij}^2) over all societies in the sample is

$$Q_{ij}^2 = \sum_{s=1}^S \frac{\Psi_{s,ij}^2}{\phi_{s,ij}} = \sum_{s=1}^S \frac{(f_{s,i} X_{s,j} - f_{s,j} X_{s,i})^2}{(X_{s,i})^2 + (X_{s,j})^2}, \quad (19)$$

which is independent of N_s as expected. However, Q_{ij}^2 is dependent on $w_{s,i}$, which is consistent with the setup of our model but different from the model in ref. 12. Thus the total sum of the squared errors is

$$Q^2 = Q_{12}^2 + Q_{13}^2 + Q_{23}^2. \quad (20)$$

One can see that a_s , b_s and c_s could be society-specific and are not included in the final objective function of model calibration.

For each pair of I_c and t , a society is excluded in model calibration if the number of agents having at least one collaborator is less than 500 to ensure that e_a has enough realizations. Changing this threshold around 500 results in same results. In addition, if the number of societies included in a model is less than 50, we do not calibrate the model because the model has 10 parameters.

To find the solution to the minimization of Q^2 , the taboo search algorithm is adopted⁴⁷. The solution space is restricted to $0 \leq \gamma_{ij} \leq 2$ for $i, j \in \{1, 2, 3\}$ and $\frac{\alpha}{2(\beta-\alpha)} > 0$. Because there are 10 free parameters, it is not easy to reach the global minimum. We thus perform a taboo search in each cell of a 9-dimensional lattice of size 2^9 with the constraint that $0 \leq \gamma_{ij} \leq 1$ or $1 \leq \gamma_{ij} \leq 2$. The parameters in certain cell corresponding to the minimum of Q^2 in all cells are obtained as the solution. The normality assumption of fitting errors has been verified by QQ-plots (Fig. S7), which rationalizes the setup of the model. We note that the partitioning of the solution space into a 9-dimensional

lattice of size 2^9 is very important. If we perform the taboo search directly, the resulting Q^2 value is significantly larger and the three preference curves $\gamma_{ij}(t)$ for each i are not well separated around $\gamma_{ij}=1$ (cf. Fig. 2, Fig. S4 and Fig. S5).

Significance tests. To test whether the preference coefficient γ_{ij} of i -agents to j -agents is significantly different from the no-preference case, we perform F -tests using the null hypothesis

$$H_0: \gamma_{ij} = 1, \quad i, j \in \{1, 2, 3\}. \quad (21)$$

Following ref. 12, the F -statistic is

$$F = \frac{\frac{SSR_{\text{con}} - SSR_{\text{uncon}}}{P_{\text{uncon}} - P_{\text{con}}}}{\frac{SSR_{\text{uncon}}}{n - P_{\text{uncon}}}}, \quad (22)$$

where SSR is the sum of squared residuals of the best-fit calibration, p is the number of model parameters, n is the number of observations, while the subscript “con” indicates the constrained model under the null hypothesis and the subscript “uncon” the unconstrained model.

Economic output of individuals. There are two virtual currencies, *Xingbi* and *Jinbi*. *Xingbi* cannot be produced by an agent’s activity and can only be bought from the system, which has an approximately stable exchange rate in reference to the Chinese currency *Renminbi*. *Xingbi* is thus a universal currency across different virtual worlds. *Jinbi*, on the other hand, is produced by the economic activities of the agents. There is a built-in exchange platform in each virtual world so that agents can exchange *Xingbi* and *Jinbi*. In this way, there is a real-time exchange rate from *Jinbi* to *Xingbi*.

An agent can produce virtual items (e.g., weapons, clothes, and medicines) and a limited amount of the virtual currency *Jinbi*. We convert the produced items and *Jinbi* to *Xingbi* to obtain the real economic output of each agent on each day. There is a marketplace in each virtual world in which agents can sell their items that are priced in *Xingbi* or *Jinbi*. The price of an item is determined by the average price of all the trades in the marketplace on a given day. Each produced item can thus be measured in *Xingbi*.

References

- Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
- Newman, M. E. J. The structure and function of complex networks. *SIAM Rev.* **45**, 167–256 (2003).
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. Complex networks: Structure and dynamics. *Phys. Rep.* **424**, 175–308 (2006).
- Fortunato, S. Community detection in graphs. *Phys. Rep.* **486**, 75–174 (2010).
- Estrada, E., Hatano, N. & Benzi, M. The physics of communicability in complex networks. *Phys. Rep.* **514**, 89–119 (2012).
- Holme, P. & Saramäki, J. Temporal networks. *Phys. Rep.* **519**, 97–125 (2012).
- Simon, H. A. On a class of skew distribution functions. *Biometrika* **42**, 425–440 (1955).
- Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
- Papadopoulos, F., Kitsak, M., Serrano, M. A., Boguñá, M. & Krioukov, D. Popularity versus similarity in growing networks. *Nature* **489**, 537–540 (2012).
- McPherson, M., Smith-Lovin, L. & Cook, J. M. Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.* **27**, 415–444 (2001).
- Curranini, S., Jackson, M. O. & Pin, P. An economic model of friendship: homophily, minorities and segregation. *Econometrica* **77**, 1003–1045 (2009).
- Curranini, S., Jackson, M. O. & Pin, P. Identifying the roles of race-based choice and chance in high school friendship network formation. *Proc. Natl. Acad. Sci. USA* **107**, 4857–4861 (2010).
- Cheng, S. W. & Xie, Y. Structural effect of size on interracial friendship. *Proc. Natl. Acad. Sci. USA* **110**, 7165–7169 (2013).
- Kovanen, L., Kaski, K., Kertész, J. & Saramäki, J. Temporal motifs reveal homophily, gender-specific patterns, and group talk in call sequences. *Proc. Natl. Acad. Sci. USA* **110**, 18070–18075 (2013).
- Coleman, J. S. Relational analysis: The study of social organizations with survey methods. *Hum. Organ.* **17**, 28–36 (1958).
- Moody, J. Race, school integration, and friendship segregation in America. *Amer. J. Soc.* **107**, 679–716 (2001).
- Kossinets, G. & Watts, D. J. Origins of homophily in an evolving social network. *Amer. J. Soc.* **115**, 405–450 (2009).
- Apicella, C. L., Marlowe, F. W., Fowler, J. H. & Christakis, N. A. Social networks and cooperation in hunter-gatherers. *Nature* **481**, 497–501 (2012).
- Curry, O. & Dunbar, R. I. M. Do birds of a feather flock together? *Hum. Nat.* **24**, 336–347 (2013).
- Rogers, E. M. *Diffusion of Innovations* 5 edn (Free Press, New York, 2003).
- Johnson, N. F. et al. Human group formation in online guilds and offline gangs driven by a common team dynamic. *Phys. Rev. E* **79**, 066117 (2009).
- Smith, A. *The Wealth of Nations* (W. Strahan and T. Cadell, London, 1776).
- Bainbridge, W. S. The scientific research potential of virtual worlds. *Science* **317**, 472–476 (2007).
- Jiang, Z.-Q., Zhou, W.-X. & Tan, Q.-Z. Online-offline activities and game-playing behaviors of avatars in a massive multiplayer online role-playing game. *EPL (Europhys. Lett.)* **88**, 48007 (2009).
- Szell, M., Lambiotte, R. & Thurner, S. Multirelational organization of large-scale social networks in an online world. *Proc. Natl. Acad. Sci. USA* **107**, 13636–13641 (2010).
- Szell, M. & Thurner, S. Measuring social dynamics in a massive multiplayer online game. *Soc. Networks* **32**, 313–329 (2010).
- Thurner, S., Szell, M. & Sinatra, R. Emergence of good conduct, scaling and Zipf laws in human behavioral sequences in an online world. *PLoS One* **7**, e29796 (2012).
- Szell, M., Sinatra, R., Petri, G., Thurner, S. & Latora, V. Understanding mobility in a social petri dish. *Sci. Rep.* **2**, 457 (2012).
- Szell, M. & Thurner, S. Social dynamics in a large-scale online game. *Adv. Complex Sys.* **15**, 1250064 (2012).
- Klimek, P. & Thurner, S. Triadic closure dynamics drives scaling laws in social multiplex networks. *New J. Phys.* **15**, 063008 (2013).
- Szell, M. & Thurner, S. How women organize social networks different from men. *Sci. Rep.* **3**, 1214 (2013).
- Fuchs, B., Sornette, D. & Thurner, S. Fractal multi-level organisation of human groups in a virtual world. *Sci. Rep.* **4**, 6526 (2014).
- Marsden, P. V. Homogeneity in confiding relations. *Soc. Networks* **10**, 57–76 (1988).
- Williams, D. The mapping principle, and a research framework for virtual worlds. *Commun. Theory* **20**, 451–470 (2010).

35. Bailenson, J. N., Blascovich, J., Beall, A. C. & Loomis, J. M. Interpersonal distance in immersive virtual environments. *Pers. Soc. Psychol. Bull.* **29**, 819–833 (2003).
36. Castranova, E. On the research value of large games. *Games and Culture* **1**, 163–186 (2006).
37. Yee, N., Bailenson, J. N., Urbaneck, M., F., C. & Merget, D. The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. *CyberPsychology & Behavior* **10**, 115–121 (2007).
38. Lofgren, E. T. & Fefferman, N. H. The untapped potential of virtual game worlds to shed light on real world epidemics. *Lancet Infectious Diseases* **7**, 625–629 (2007).
39. Grabowski, A. & Kruszewska, N. Experimental study of the structure of a social network and human dynamics in a virtual society. *Int. J. Mod. Phys. C* **18**, 1527–1535 (2007).
40. Xie, W.-J., Li, M.-X., Jiang, Z.-Q. & Zhou, W.-X. Triadic motifs in the dependence networks of virtual societies. *Sci. Rep.* **4**, 5244 (2014).
41. Bhattacharya, H. & Dugar, S. Partnership formation: The role of social status. *Manag. Sci.* **60**, 1130–1147 (2014).
42. Ball, B. & Newman, M. E. J. Friendship networks and social status. *Network Science* **1**, 16–30 (2013).
43. Dunbar, R. I. M., Arnaboldi, V., Conti, M. & Passarella, A. The structure of online social networks mirrors those in the offline world. *Soc. Networks* **43**, 39–47 (2015).
44. Eagle, N., Macy, M. & Claxton, R. Network diversity and economic development. *Science* **328**, 1029–1031 (2010).
45. Bettencourt, L. M. A., Samaniego, H. & Youn, H. Professional diversity and the productivity of cities. *Sci. Rep.* **4**, 5393 (2014).
46. Ortman, S. G., Cabaniss, A. H. F., Sturm, J. O. & Bettencourt, L. M. A. Settlement scaling and increasing returns in an ancient society. *Sci. Adv.* **1**, e1400066 (2015).
47. Cvijović, D. & Klinowski, J. Taboo search: An approach to the multiple minima problem. *Science* **267**, 664–666 (1995).

Acknowledgements

W.J.X., M.X.L., Z.Q.J. and W.X.Z. received support from the National Natural Science Foundation of China (NSFC) Grants 11505063, 11375064 and 11205057, and the Ph.D. Programs Foundation of Ministry of Education of China Grant 20120074120028. B.P. and H.E.S. received support from the Defense Threat Reduction Agency (DTRA), the Office of Naval Research (ONR), and the National Science Foundation (NSF) Grant CMMI 1125290.

Author Contributions

W.-J.X., B.P., W.-X.Z. and H.E.S. conceived the research. W.-J.X. and W.-X.Z. designed the analyses. W.-J.X., M.-X.L. and Z.-Q.J. compiled the data. W.-J.X., M.-X.L., Z.-Q.J. and W.-X.Z. conducted the analyses. All authors discussed the results. W.-J.X., B.P., W.-X.Z. and H.E.S. wrote the paper.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Xie, W.-J. *et al.* Skill complementarity enhances heterophily in collaboration networks. *Sci. Rep.* **6**, 18727; doi: 10.1038/srep18727 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>