Eur. Phys. J. B **58**, 97–105 (2007)
DOI: 10.1140/epjb/e2007-00200-x

**THE EUROPEAN PHYSICAL JOURNAL B**

# An evolutionary model of social networks

M. Ludwig[a] and P. Abell

Department of Management, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK

Received 30 January 2007 / Received in final form 11 May 2007
Published online 28 July 2007 – 

**Abstract.** Social networks in communities, markets, and societies self-organise through the interactions of many individuals. In this paper we use a well-known mechanism of social interactions — the balance of sentiment in triadic relations — to describe the development of social networks. Our model contrasts with many existing network models, in that people not only establish but also break up relations whilst the network evolves. The procedure generates several interesting network features such as a variety of degree distributions and degree correlations. The resulting network converges under certain conditions to a steady critical state where temporal disruptions in triangles follow a power-law distribution.

**PACS.** 89.75.-k Complex systems – 87.23.Ge Dynamics of social systems

## 1 Introduction

Complex systems like organisations and organisms often take the form of networks — sets of actors, cells or other units tied together by edges. To explain the behaviour of complex systems, network models have been developed in many fields, for example, in physics, biology, operational research, economics, and sociology [1,2]. Most of these models reproduce the observed properties of biological and technical networks well but provide less accurate descriptions of social networks. The reason for this could be that people — unlike cells or particles — pursue individual goals that are mostly responsible for their social contacts [3]. These goals affect the network but are also affected by the network [4]. The goals as well as the network are not static but co-evolve over time.

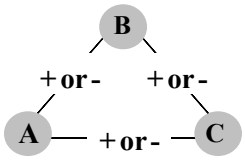
Of course, several general principles of network construction could apply in the social as well as in the physical domain [1]. Take, for example, three fairly common growth principles of networks: *random attachment*, *preferential attachment, and age-driven removal*. *Random attachments* might happen within groups of people with no previous contacts at all (say, on a cruise). *Preferential attachment* could be at work when people with a larger number of friends tend to acquire new friends more readily (evoking the *Matthew effect*: "For to every one that hath shall be given"). *Age-driven removals* take place as people die or fall into oblivion (for example, forgotten High School friends). These principles surely play a role in the emergence of social networks; however, they only describe wholesale phenomena, insensitive to individuals' goals. For example, a person might choose to contact a less popular person if his position in the social network makes it preferential to do so.

How then can we extract a feasible construction principle out of the myriad of individual goals in social groups? Sociology offers at least two findings: the *locality principle* and *structural balance*. The locality principle describes the fact that people mostly choose their social contacts based on their local information of the network [3]. For example, people might become acquainted with each other through the introduction by a common friend. Using this argument, [5] simulated the evolution of social networks by randomly linking up neighbouring nodes. Such an introduction mechanism alone, however, does not take into account two other cornerstones of social life: the quality of dyadic relations (Do two people like/dislike each other?) and triadic relations (Do two people compete for the attention, co-operation, etc. of the third person?). These two aspects of social interactions are major drivers of social choice and at the heart of another classic in Sociology, *Structural Balance Theory*.

*Structural Balance Theory* evolved from the work of [6] and describes a social selection process in people's minds. According to this theory, people establish dyadic relations that each side equivalently perceives as either positive or negative. If three persons form a triadic relation they perceive it as either "balanced" or "imbalanced", depending on the number of positive and negative relations in the triangle (see Fig. 1).

A balanced triangle exists if either one or all of the three relations are positive, that is, if "*my friend's friend is my friend*", or "*my enemy's enemy is my friend*", or

[a] e-mail: m.ludwig@lse.ac.uk




| Triangle type | (A;B) | (A;C) | (B;C) |
|---|---|---|---|
| Balanced | + | + | + |
| | + | - | - |
| | - | + | - |
| | - | - | + |
| Imbalanced | - | + | + |
| | + | - | + |
| | + | + | - |
| | - | - | - |

**Fig. 1.** Positive (+) and negative (-) sentiments in triangle relations and the respective triangle type.

"*my enemy's friend is my enemy*", or "*my friend's enemy is my enemy*". An imbalanced triangle, in contrast, occurs in all other combinations, that is if either two or none of the three relations are positive. Imbalanced triangles provoke unease and force people towards more balanced combinations that could involve a re-organisation of the entire network [7]. These effects of triadic relations have been confirmed in several empirical studies [8,9]. It, thus, seems reasonable to use the balance of sentiment in triadic relations as a construction principle for social networks.

In the following we introduce a network model that is based on the insights of balance theory as well as the locality principle. To model the network growth accordingly, we sequentially randomly attach positive and negative edges to a given set of nodes. For each node, we sequentially keep track of the number of unbalanced triangles. Once a node reaches a certain threshold of unbalanced sentiments, we remove its links at random one after the other until the threshold is not exceeded. This process in turn might cause other nodes to become too imbalanced so that the re-balancing process cascades until all affected nodes are sufficiently balanced again.

We first show how to model such a network's evolution and then analyse how the evolutionary process converges towards an unstable equilibrium that may be a state of self-organised criticality. The concept of self-organised criticality, first outlined by [10] in the "sand pile" model, increasingly stimulates research into the description and construction of networks (see, among others, [11–15]). For example, [12] use a network mechanism to describe self-organised criticality in the sun's magnetic field lines and the resulting size distribution of solar flares. [15] show how a random network evolves into a state of self-organised criticality upon introducing a rewiring procedure that depends on each link's age. Although these models make reference to the original sand pile idea, they usually apply different definitions of self-organised criticality in networks. We here define a network's self-organised criticality as a medium-term statistical steady state where the average number of added and removed links (or triangles) per time step is equal, and the distribution of triangle removals in a given time unit is scale-invariant [16,17]. The properties of networks in such a state will be reported later in the article.

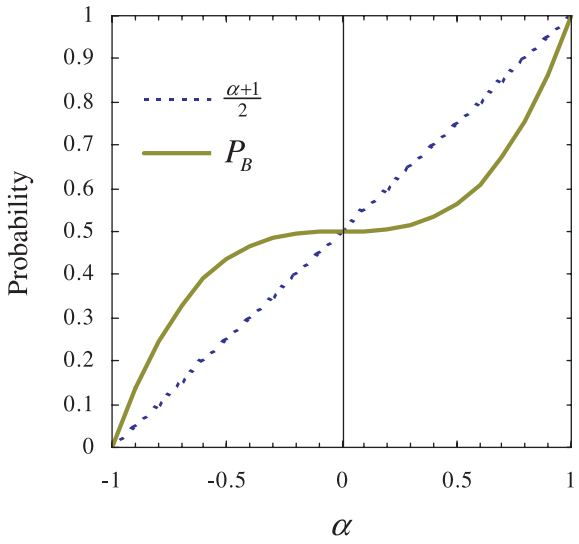



**Fig. 2.** Probabilities for positive links and balanced triangles in the network for different values of $\alpha$. If $\alpha = 0$, both probabilities are 50%.

## 2 Model of network evolution

Consider a set of $n$ vertices (that is, persons) that are subsequently linked with each other. At each time step $t$ a single symmetric positive or negative link is established at random between two directly unconnected vertices. The likelihood of a link's quality (positive or negative) depends on a *friendliness index* $-1 \leqslant \alpha \leqslant 1$ so that the probability of a positive link is $\frac{\alpha+1}{2}$. Accordingly (see Fig. 1), the probability of a randomly chosen completed triangle being balanced (respectively unbalanced) is

$$P_B = \binom{3}{3}\left(\frac{\alpha+1}{2}\right)^3 + \binom{3}{1}\left(1-\left(\frac{\alpha+1}{2}\right)\right)^2\left(\frac{\alpha+1}{2}\right) = \frac{\alpha^3+1}{2};$$
$$P_U = 1 - \frac{\alpha^3+1}{2} = \frac{1-\alpha^3}{2}. \quad (1)$$

If we have, for example, $\alpha = 0.4$, then 70% of newly introduced links are positive and $\frac{1}{2}\left(0.4^3+1\right) = 53.2\%$ of triangles are balanced (see Fig. 2). In general, the fraction of balanced triangles in the network shifts only slightly as the friendliness index varies between $-0.5 \leqslant \alpha \leqslant 0.5$. It is only for extreme values for $\alpha$ that the mix of triangles changes dramatically.

Now consider a uniformly distributed threshold parameter $-1 \leqslant \beta \leqslant 1$ that indicates the quantity of imbalanced triangles that is just tolerated by an agent. We compare $\beta$ with a vertex $i$'s balance index $\varphi_i = \left(\frac{\triangle^+ - \triangle^-}{\triangle^+ + \triangle^-}\right)_i$, where $\triangle^+$ and $\triangle^-$ are the number of balanced and imbalanced triangles running through the vertex. Hence, a vertex $i$ stays inert as long as $\beta \leqslant \varphi_i$, and becomes unbalanced otherwise. After a new link is added to the network, all $\varphi_i$ of the network are calculated in a random sequence. If a

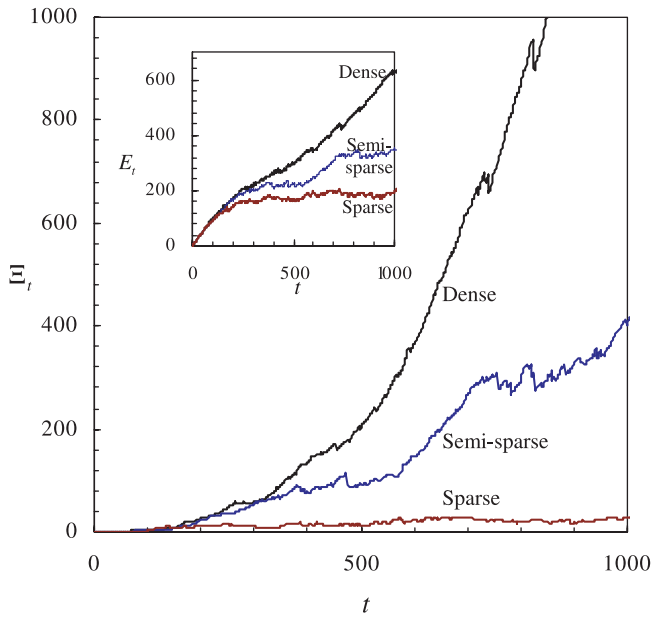



**Fig. 3.** Typical growth patterns of triangles $\Xi_t$ and links $E_t$ in dense, semi-sparse, and sparse networks ($n = 60$) during the early time steps of the evolution.

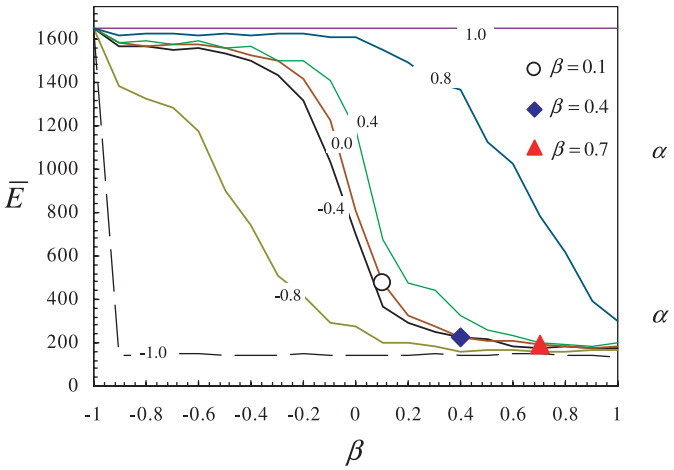



**Fig. 4.** Average number of links $\bar{E}$ for various combinations of $\alpha$ and $\beta$; each line represents a different friendliness index $\alpha = \{-1.0; -0.8; -0.4; 0.0; 0.4; 0.8; 1.0\}$. Data are averaged results over $1600 < t \leqslant 1700$ and 4 simulations. The three highlighted combinations $\beta = \{0.1; 0.4; 0.7\}$ lie on the line of $\alpha = 0$ and are analysed in more detail later in the paper.

vertex $i$ falls below the threshold $\beta$, one of its links is randomly removed and its balance index $\varphi_i$ is recalculated. If this vertex is still unbalanced the procedure continues until it is balanced again. Now all remaining $\varphi_i$ are again determined in the same sequence as before, and the procedure goes on for each unbalanced vertex until all vertices are sufficiently balanced again. Only then a new link is attached to the network.

It could be argued that an agent does not randomly remove one of his links but rather tries to abandon the link that adds most to the unbalanced condition. To do this, however, the agent has to closely keep track of each link's triangle contributions, which seems to be a daunting, if not impossible task for networks of any complexity. Hence, random link removals are quite plausible, especially if they are interpreted as an actor's attempt to become more restrained in general.

The network's evolution proceeds either until the network contains the maximum number of potential links $n(n-1)/2$ [18], or until a predetermined number of time steps is reached. For each time step, we measure several network properties, the number of link removals as well as number and type of triangle removals.

## 3 Settings for the network's evolution

Three types of networks occur in the simulations: *dense*, *semi-sparse*, and *sparse networks* (see Fig. 3). In dense networks the number of triangles $\Xi_t$ and links $E_t$ at time step $t$ quickly grows, interrupted by little cascades of break-ups, until the network becomes complete at or soon after $t = n(n-1)/2$ is reached. Sparse networks hardly have any triangles and accumulate only a relative small number of links up to a certain level around which both $\Xi_t$ and $E_t$ fluctuate. Semi-sparse networks have a similar growth pattern to sparse networks; however, the level around which their number of triangles and links fluctuate is significantly higher than in sparse networks.

The type of network depends on both the friendliness index and the (assumed) uniform balance threshold in the network. To investigate the "space of networks" we vary the friendliness index and the balance threshold for representative values. For each combination, the network size is $n = 60$ and the duration of the evolution is $t_{\max} = 1700$, chosen to be well before the time $t = 60 \times 59/2 = 1770$ when the network could become complete. We measure the average number of links between $1600 < t \leqslant t_{\max}$ and repeat this procedure 4 times to calculate the average number of edges $\bar{E}$ over all four network evolutions (see Fig. 4).

For each friendliness index $\alpha$ we obtain a different function between the balance threshold and the average number of links. For $\alpha = 1.0$, the network evolves to the maximum number of 1700 links regardless of the balance threshold as a new link is added at each time step and no break-ups occur. This case is thus equivalent to the classical random graph model (also called ER-model) [19]. The same ER-network is in place for $\beta = -1$, when the nodes tolerate an unlimited number of negative triads.

If $\alpha = -1.0$, the network contains 1700 links when $\beta = -1$, but only about 145 links for most other balance thresholds. A special combination is $\alpha = -1.0$ and $\beta = 1$ where no triangles exist at all and the network only features tree graphs. For all other values of $\alpha$, we find similar functions between $\beta$ and $\bar{E}$, each featuring dense, semi-sparse, and sparse networks.

If $\beta$ is sufficiently low, the evolved network is dense and contains only slightly less than 1700 links. As $\beta$ becomes more positive, the number of links strongly decreases and

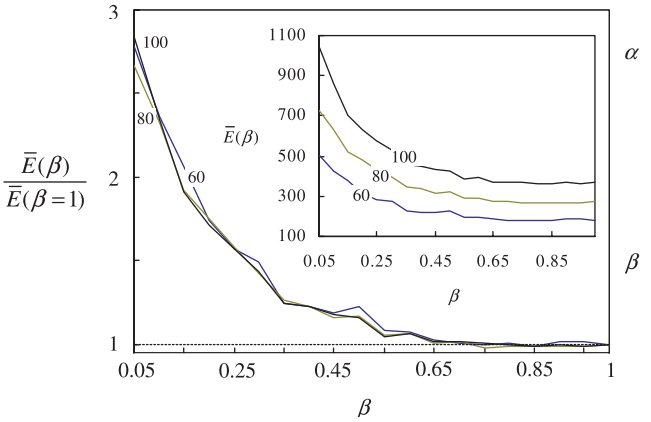



**Fig. 5.** The average number of links $\bar{E}$ for $\beta = \{0.05, 0.10, ..., 0.95, 1\}$ and $\alpha = 0$ standardised by $\bar{E}(\beta = 1)$. The number of links were averaged between $1000 \leqslant t \leqslant n(n-1)/2$ and, respectively, over 5 runs for the three network sizes $n = \{60, 80, 100\}$. The function $\bar{E}(\beta)/\bar{E}(\beta = 1)$ is almost identical for the three cases. At $\beta \approx 0.65$, the function becomes significantly larger than 1 and strongly increase as $\beta$ tends to zero. Inset: the number of links $\bar{E}(\beta)$ for the same set-up. The slope of $\bar{E}(\beta)$ seems to be independent of $n$ for sufficiently large $\beta$.

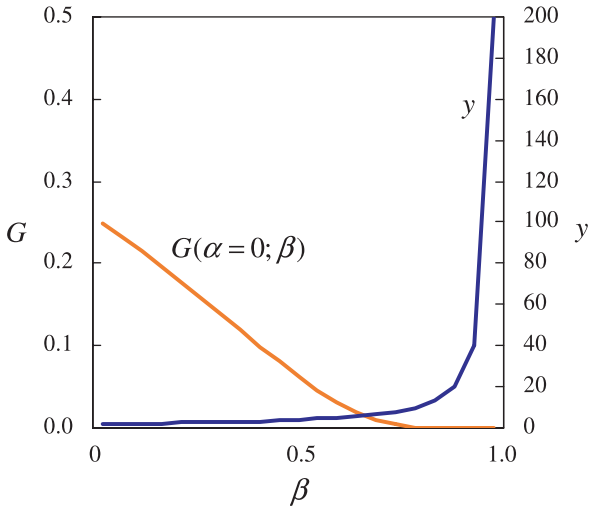



**Fig. 6.** The probability $G(\alpha = 0; \beta)$ that a randomly chosen set of $y$ triangles ever created throughout the evolution contains only balanced triangles. For sufficiently high values of $\beta$ (here, about $\beta = 0.75$), $y$ increases very fast and converges to infinity for $\beta = 1$. Consequently, $G(\alpha; \beta)$ is approximately zero for values $\beta > 0.75$ (and exactly zero for $\beta = 1$). As a result, the number of triangles and links in the network falls strongly at around $\beta = 0.75$ and stays at about the same low level for higher values of $\beta$. This cut-off value increases for larger network sizes $n$ and longer durations of the evolution, $t_{\max}$.

the networks become semi-sparse. Beyond a certain degree of intolerance (balance threshold), the number of links is very low and decreases only slightly so that only sparse networks evolve.

We can check for network size effects by running simulations similar to those conducted above for parameter combinations with $\alpha = 0, \beta = \{0.05, 0.10, ..., 0.95, 1\}$, and $n = \{60, 80, 100\}$. The insert of Figure 5 displays $\bar{E}$ calculated as the mean number of links between $1000 \leqslant t \leqslant n\,(n-1)/2$, averaged over 5 runs. The average link number grows as the network size increases, however, the slope of $\bar{E}$ as a function of $\beta$ seems to be independent of the network size for sufficiently high threshold values (here, $\beta > 0.05$). We can substantiate this claim if we divide $\bar{E}$ by $\bar{E}(\beta = 1)$, as shown in Figure 5. Of course, if the threshold level $\beta$ is sufficiently low (here, $\beta < 0.05$), the slope of $\bar{E}(\beta)$ becomes steeper the larger is the network.

To analyse what drives the boundaries between the three types of networks, let us first take into account two measures: the probability for a balanced triangle $P_B = \frac{1+\alpha^3}{2}$ and the minimum proportion of balanced triangles required by the network's members $\frac{1+\beta}{2}$ to remain inert. The boundary between dense and semi-sparse networks appears to take place if the probability for a balanced triangle is lower than the required minimum proportion of balanced triangles:

$$P_B < \frac{\beta+1}{2} \Rightarrow \frac{1+\alpha^3}{2} < \frac{1+\beta}{2} \Rightarrow \alpha^3 < \beta. \qquad (2)$$

Otherwise the probability $U$ that a node retains an unbalanced triangle at time step $t$ quickly increases from $U = P_U = \frac{1-\alpha^3}{2}$ at $t = 0$ to $U = 1$ so that the network becomes dense.

If $\alpha^3 < \beta$, the probability $U$ is still a function of $P_U$ but also depends on the required minimum proportion of balanced triangles $\frac{\beta+1}{2}$. The latter translates into a required number, $y$, of balanced triangles that a node has to accumulate before it retains its next unbalanced triangle (unless $\beta = -1$, so that all triangles are tolerated). For this number $y$, it must hold that

$$\frac{y-1}{y} = \frac{\beta+1}{2} \Leftrightarrow y = \frac{2}{1-\beta}. \qquad (3)$$

We then can say that the probability $U$ also depends on the function $G(\alpha; \beta)$

$$G(\alpha;\beta) = \left(\frac{1+\alpha^3}{2}\right)^{y} = \left(\frac{1+\alpha^3}{2}\right)^{2/(1-\beta)}, \qquad (4)$$

which is the probability that a randomly chosen set of $y$ triangles created at any time throughout the evolution, contains only balanced triangles. If $G(\alpha; \beta)$ is significantly above zero, the average number of edges $\bar{E}$ in the network strongly increases and the network is likely to become semi-sparse. In Figure 6, we plot $G$ for $\alpha = 0$ and $y$ against different values of $\beta$.

Apparently, $G$ decreases steadily as $\beta$ grows. At a certain value of $\beta$ (for $\alpha = 0$, at about$\beta = 0.75$), the probability $G$ is close to zero. This follows from the fact that

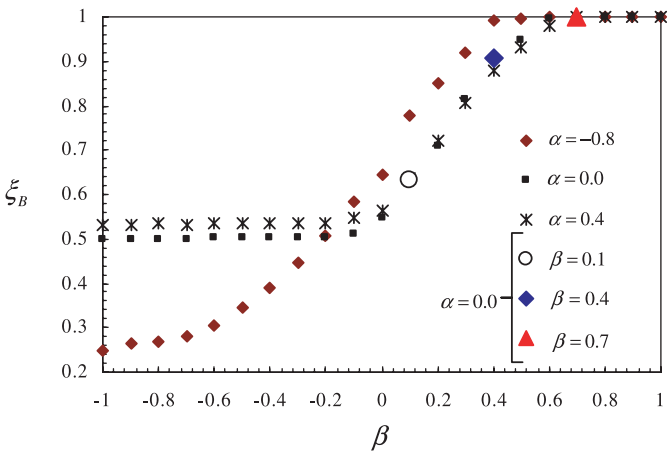



**Fig. 7.** The fraction of balanced triangles $\xi_B$ for different values $\beta$ and for $\alpha = \{-0.8; 0; 0.4\}$. For values $\alpha^3 \geqslant \beta$, the fraction of balanced triangles $\xi_B$ is close to $\frac{1}{2}\left(1+\alpha^3\right)$. Beyond that point, $\xi_B$ steadily increases until it is close to one. The usual examples $\beta = \{0.1; 0.4; 0.7\}$ are for $\alpha = 0$.

$y$ increases dramatically beyond $\beta = 0.75$, and we find sparse networks.

However, the exact boundary condition between sparse and semi-sparse networks not only depends on $G$ but also on the network size $n$, the duration $t_{\max}$ of the network's evolution, and other random effects during the evolution.

To gain a better idea of the three types of networks, we measure the number of balanced and unbalanced triangles in the network at each time step $t$ ($\Xi_t^+$ and $\Xi_t^-$), and calculate the average proportion of balanced triangles $\xi_B = \frac{1}{t_{\max}-t_{\min}} \sum_t \frac{\Xi_t^+}{\Xi_t^+ + \Xi_t^-}$ for a given duration $t_{\min} < t \leqslant t_{\max}$. The graph in Figure 7 shows $\xi_B$ for $\alpha = \{-0.8; 0; 0.4\}$ (simulated for $n = 60$ and averaged over 100 time steps between $1600 < t \leqslant 1700$ and over 30 simulations) against different values of $\beta$.

For lower values of $\beta$, the fraction of unbalanced triangles is close to (or exactly, for $\beta = -1$) $\frac{1}{2}\left(1+\alpha^3\right)$ as long as $\alpha^3 \geqslant \beta$. In this phase almost all unbalanced triangles are accepted after a sufficient number of time steps so that $\xi_B \approx P_B$. If $\alpha^3 < \beta$, the proportion of balanced triangles goes up as $\beta$ increases, and approaches $\xi_B = 1$ for values of $\beta = 0.4$ to $\beta = 0.6$. For example, the three cases $\beta = \{0.1; 0.4; 0.7\}$ for $\alpha = 0$ have a fraction of balanced triangles of about 65%, 90%, and 100%. Interestingly, the fraction of balanced triangles in semi-sparse networks are *higher* for cases of lower values of $\alpha$. Moreover, it becomes clear that a proportion of balanced triangles $\xi_B \approx 1$ corresponds to sparse networks.

The fact that a very high fraction of balanced triangles indicates a sparse network allows us to use it as an order parameter for identifying the boundary between semi-sparse and sparse networks.

Let us define the margin $\varepsilon = 1 - \xi_B \approx 0$ by which the proportion of balanced triangles in the network is lower than 1. This margin is driven by $G(\alpha;\beta)$, the network size $n$, the duration of the evolution $t_{\max}$ and other random effects during the network's evolution. If we set everything else constant, we can define a sparse network in terms of $\varepsilon$ and $G(\alpha;\beta)$ where the following holds

$$G(\alpha;\beta) \leqslant \varepsilon \Leftrightarrow \frac{\beta+1}{2} \leqslant \frac{\log(\varepsilon)}{\log\left(\frac{1+\alpha^3}{2}\right)} \Leftrightarrow$$

$$\beta \leqslant 1 - \frac{2\log\left(\frac{1+\alpha^3}{2}\right)}{\log(\varepsilon)}. \quad (5)$$

All networks whose fraction of balanced triangles is $\xi_B \geqslant 1-\varepsilon$ are thus defined as sparse. This allows us to indicate the boundary condition between semi-sparse and sparse networks for different combinations of $\alpha$ and $\beta$. To this end, we simulate networks with $n = 100$ and measure the fraction of balanced triangles for $10\,000 < t \leqslant 30\,000$. If a network becomes complete during this simulation, it is classified as dense. Accordingly, a sparse network is a non-complete network whose fraction of unbalanced triangles falls below a given level $\varepsilon$. This level increases as the evolution's duration increases. For the example's duration $t_{\max} = 30\,000$, the sparse networks appear to occur for approximately $\varepsilon = 0.03$. Thus non-complete networks with $\xi_B < 0.97$ are labelled as semi-sparse. In Figure 8a, we give an overview of this classification for combinations of $\alpha = \{-1, -0.8, ..., 0.8, 1\}$ and $\{-1, -0.8, ..., 0.8, 1\}$. Apparently, the area of semi-sparse networks lies between the sparse and dense networks in the $\alpha - \beta$-space. We can now compare the simulation results with the two boundary conditions stated in (2) and (5) (see Fig. 8b). Using $\varepsilon = 0.03$ in (5), we can closely reproduce the shape and location of the boundary between sparse and semi-sparse networks. The example networks for $\alpha = 0, \beta = \{0.1; 0.4\}$ are semi-sparse, whilst the network for $\alpha = 0, \beta = 0.7$ is sparse. We also show the boundary between dense and semi-sparse networks as given by (2). At $t_{\max} = 30\,000$, all networks with $\alpha^3 > \beta$ are complete whilst most other networks have fewer links. However, there are also complete networks for $\alpha^3 \leqslant \beta$ (for example, in case of $\alpha = 0.8, \beta = 0.8$). So obviously, the boundary conditions are not entirely fixed. As we will see in the next section, the links and triangles of networks in and around the semi-sparse area fluctuate considerably during the evolution. This can result in sparse networks becoming semi-sparse and semi-sparse networks becoming dense.

## 4 Simulating the network's evolution

The network's evolution is interesting in two respects: first, the number of break-ups that occur during each time step of its evolution, and second, the corresponding development of network traits. To describe a network, we calculate the following properties after each time step $t$: $E_t$, $\Xi_t^+$ and $\Xi_t^-$, the proportion of balanced triangles $\xi_{B,t} = \frac{\Xi_t^+}{\Xi_t^+ + \Xi_t^-}$, the number of newly formed triangles $d\Xi_t$ at the beginning of time step $t$, the break-ups of the number of links and triangles ($\partial E_t = |E_t + 1 - E_{t-1}|$ and $\partial \Xi_t = |\Xi_t + d\Xi_t - \Xi_{t-1}|$), the degree distribution $p_t(k)$,

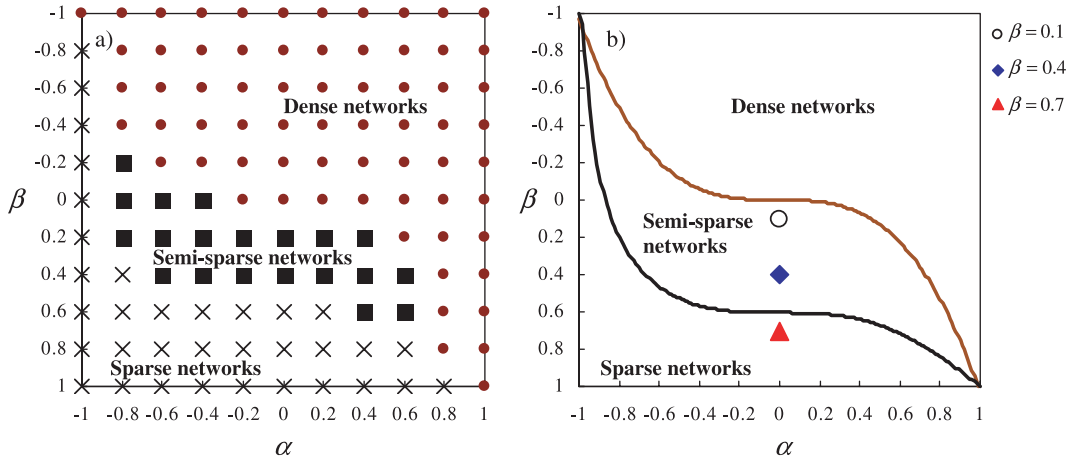



**Fig. 8.** (a) Schematic location of sparse, semi-sparse, and dense networks in the $\alpha - \beta$-space. The networks are of size $n = 100$ and are analysed for the duration of $10\,000 < t \leqslant 30\,000$. Completed networks are classified as dense. Non-complete networks whose share of balance triangles is $\xi_B > 1 - \varepsilon$ are classified as sparse, while all other non-complete networks are labelled as semi-sparse. (b) The boundary conditions $\beta = \alpha^3$ and $\beta = 1 - 2\log\left(\frac{1+\alpha^3}{2}\right)/\log(\varepsilon)$ with $\varepsilon = 0.03$ almost completely envelope the area of semi-sparse networks found in the simulation, suggesting that both conditions are major drivers of the change between network types. Some completed networks can occur below the upper boundary conditions, which is discussed below. For our standard examples with $\alpha = 0$, we expect a sparse network for $\beta = 0.7$ and semi-sparse networks for $\beta = \{0.1, 0.4\}$.

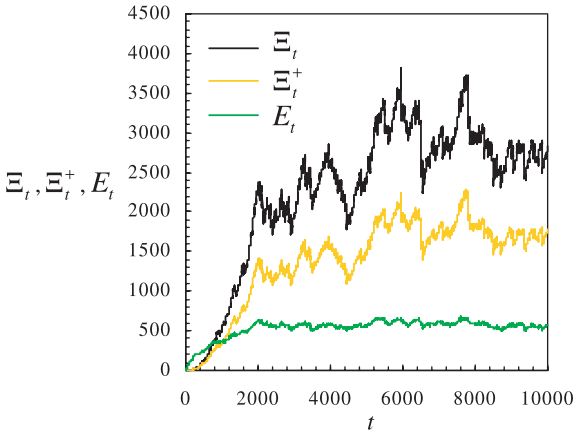



**Fig. 9.** The evolution of $E_t$, $\Xi_t$, and $\Xi_t^+$ during 10 000 time steps for $\alpha = 0; \beta = 0.1$. The number of links is in a state of steady criticality after about 1200 time steps, while the number of triangles and positive triangles reaches the same state some time later. When the network becomes steady critical, the changes in triangles and links can be considerable (see, for example, the period shortly before $t = 6500$.

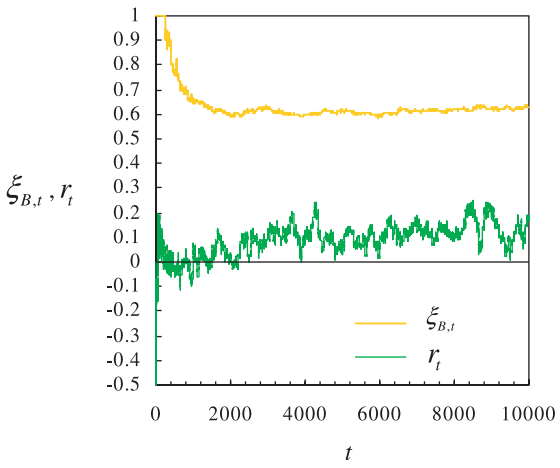



**Fig. 10.** The evolution of $\xi_{B,t}$ and $r_t$ during 10 000 time steps for $\alpha = 0; \beta = 0.1$. The fraction of positive triangles $\xi_{B,t}$ decrease from 1 to a stationary value (here, about 0.6). The degree correlation first assumes negative values before turning positive and fluctuating around a positive value.

and the degree correlation $r_t$. The degree distribution indicates the fraction $p(k)$ of vertices with $k$ links in a network. A positive degree correlation $r$ represents the tendency of network nodes to be connected to nodes of similar degree. It is defined as the Pearson correlation coefficient of degrees at either end of a link and assumes values between $-1 \leqslant r \leqslant 1$ [20].

A network's evolution always passes through a starting or "build up" phase, as shown in Figure 9. The plot depicts the number of links, triangles, and positive triangles during the evolution of a network with $\alpha = 0$ and $\beta = 0.1$. After the start-up phase is finished (here, after about 1600 time steps), the number of links starts fluctuating around a mean value (in this setting, at about 33% of the maximum number of potential links $60 \cdot 59/2 = 1770$). The fluctuating (or "stationary") state begins earlier the higher is the balance threshold $\beta$. However, its start can be difficult to spot, especially for the number of triangles. As observable in Figure 9, the number of triangles approaches the fluctuating state much later and is more volatile than the number of links. The swings can be massive even for the short period of 10 000 time steps. In our example, there is a decrease of about 30% of triangles and of about 15% of links between time steps 6467 and 6480. The positive triangles' trajectory mostly moves in parallel to the development of all triangles. However, the

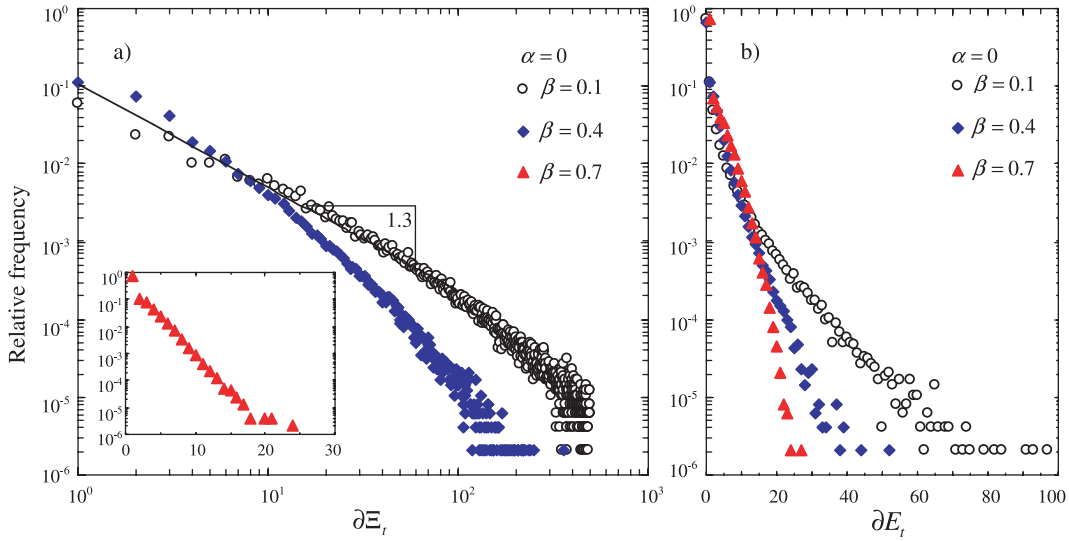



**Fig. 11.** The relative frequency distribution of (a) the number of break-ups of triangles $\partial\Xi_t$ and (b) the number of break-ups of links $\partial E_t$ in the network during $t = 10\,000$ until $t = 100\,000$ for $\beta = \{0.1; 0.4; 0.7\}$. The break-up distribution of triangles follow a power-law for $\beta = 0.1$ and $\beta = 0.4$ and an exponential distribution for $\beta = 0.7$. The break-up distribution for links seem to follow exponential functions but can also be fitted to power-law functions. Simulations with much larger networks are required to confirm this judgement. The regressions become less accurate for highly infrequent data points due to the finite size of the network and of the network's evolution.

proportion of positive triangles decreases until it converges to about 60% of total possible (Fig. 10).

The reason for this decrease is that, as the evolution of the network proceeds, more and more links and triangles are located with nodes that by chance have enjoyed a stream of predominantly balanced triangles. These "super-balanced" nodes act as a stabilising "buffer", increasing the network's capacity to absorb negative triangles.

To determine the distribution of break-ups per time step, we need to strike a balance between the network's size $n$ and the duration $t_{\max}$ of the network evolution. While the statistic of break-ups clearly requires large networks to be meaningful, the running times usually become unacceptable for very large networks. However, we can partly capture the behaviour of large networks by extending the duration of the evolution (for example, to collect more extreme outliers of break-up sizes). For these reasons we ran the network evolutions for 100 000 time steps with a relatively small network size $n = 60$. We collect data for $t > 10\,000$ in order to measure the break-ups only during the fluctuating state. As before, we choose the three standard combinations with balance thresholds of $\beta = \{0.1; 0.4; 0.7\}$ and a friendliness index of $\alpha = 0$.

Figures 11a and 11b depict respectively the relative frequency of break-ups of triangles and links during a time step $t$. The break-up distribution for triangles can be fitted to a power-law for semi-dense networks (here: $\beta = 0.1$ and $\beta = 0.4$). For example, the power-law exponent of the break-up distribution for $\beta = 0.1$ is about 1.3. As $\beta$ increases, the power-law exponent becomes larger, that is, the power-law distributions become steeper. For sparse networks (here: $\beta = 0.7$), the distribution is exponential, as shown in the semi-log plot of the insert in 11a. The break-up distributions for links can be fitted to an exponential function [21] whose mean is (very close to) 1 as long as the network is non-dense (here, for all three cases $\beta = \{0.1; 0.4; 0.7\}$). In other words, the creation of links equals the average destruction in non-complete networks with $\alpha^3 \leqslant \beta$.

According to our definition, this indicates that semi-sparse networks approach a state of self-organised criticality. Simulations of other settings show that break-ups in dense networks hardly occur while break-ups in sparse networks are frequent but of limited size. The reason for the latter is that unbalanced triangles beyond the lower boundary are almost always torn apart upon their creation, which leaves no room for the creation of network structures large enough to provoke break-ups at significant scale. So it is only in semi-sparse networks, we find that the number of triangle break-ups follows a power-law. This condition, of course, only holds if the average number of added and removed links is in equilibrium.

The degree correlation and degree distribution fluctuate throughout the evolutionary process and each is quite different in the "start up" phase and the stationary phase. Therefore, we take averages over a period of time steps that surely take place in the stationary phase and mark mean values by a bar over the respective symbol: the average degree correlation is $\bar{r} = \frac{1}{t_{\max}-t_{\min}} \sum_t r_t$ and the average degree probability is $\bar{p}(k) = \frac{1}{t_{\max}-t_{\min}} \sum_t p_t(k)$. Proceeding in this way, we can compare values for different combinations of $\alpha$ and $\beta$.

The degree correlation $\bar{r}$ assumes positive values in semi-sparse networks (see Fig. 12, using settings as in

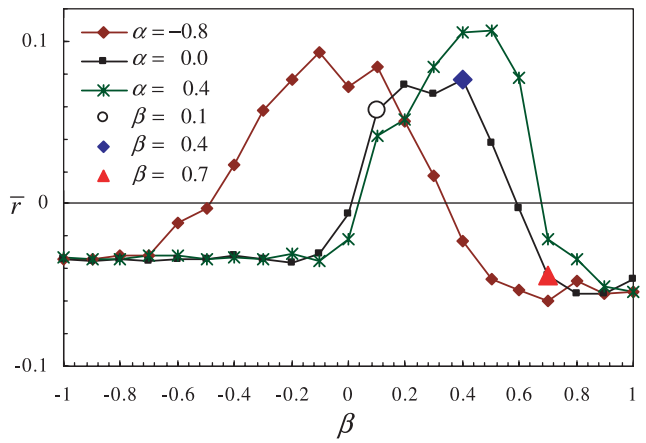



**Fig. 12.** Average degree correlation $\bar{r}$ for different values of $\beta$ (settings as in Fig. 7). Until the boundary condition at $\alpha^3 = \beta$, the degree correlation is slightly below zero (but converges to zero for larger networks as in the classical random graph). Between the first and second phase boundary condition, the network displays self-organised criticality and the degree correlation increases until about $\bar{r} = 0.1$ (and beyond 0.25 if the evolution lasts longer). Beyond the lower boundary condition, the degree correlation is slightly negative. The example combinations $\alpha = 0, \beta = \{0.1; 0.4\}$ have a positive $\bar{r}$, while the combination $\alpha = 0, \beta = 0.7$ leads to a network with a negative $\bar{r}$.

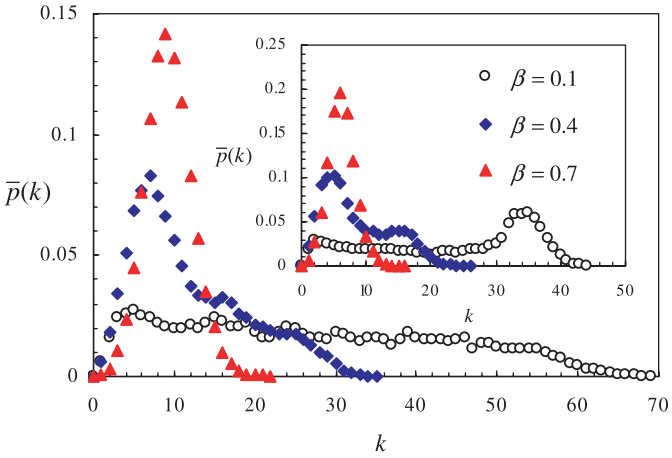



**Fig. 13.** Degree distribution $\bar{p}(k)$ (averaged over time steps between $t = 12\,000$ and $t = 20\,000$) for $\beta = \{0.1; 0.4; 0.7\}$ and a network size of $n = 200$. Inset: degree distribution $\bar{p}(k)$ (averaged over time steps between $t = 3000$ and $t = 10\,000$) for $\beta = \{0.1; 0.4; 0.7\}$ and a network size $n = 60$.

Fig. 7). The positive degree correlation is due to the fact that links and triangles gravitate to the "super-balanced" nodes. These nodes are less likely to remove links, and thus, on average, tend to be linked to each other, resulting in a positive degree correlation $r_t$ ("hard core effect"). The "hard core effect" increases as the balance threshold increases but disappears after the lower boundary condition is encountered. The reason for this on-off phenomenon of degree correlations is that an increasing balance threshold, on the one hand, makes the group of "super-balanced" nodes more exclusive and fosters links to and between them. On the other hand, it increases the number of nodes that frequently break up links and "re-randomise" the network. Both counter-steering trends determine the size of the "hard core". If the "re-randomisation" dominates, as in sparse networks, the degree correlation becomes slightly negative. In case of completed networks, the degree correlation is very close to zero, but might slightly diverge from zero if the network size is relatively small or if a sufficient number of break-ups have taken place before the network is completed.

The balance threshold $\beta$ not only has a strong impact on the degree correlation but also on the degree distribution $\bar{p}(k)$. The inset of Figure 13 shows the degree distribution $\bar{p}(k)$ for the three cases of $\beta = \{0.1; 0.4; 0.7\}$ with a network size $n = 60$. Again, the "hard core effect" is at work: if the balance threshold increases, the "super-balanced" nodes gain additional links during the network's evolution, which leads to more varied degrees and thicker right tails of the distribution. At a certain point, the "hard core effect" becomes smaller as the "re-randomisation" intensifies. If the balance threshold is high enough ($\beta = 0.7$ in the example), the degree distribution seems to converge to a Poisson distribution, as in a pure-random graph.

The average degree moves in parallel to the number of links as shown and reaches a stationary value after an initial build-up. This stationary value decreases for higher balance thresholds (see Fig. 4). All degree distributions have their mode at rather small degrees and are skewed to the right. For some combinations, (see for example, $\beta = 0.1$, $n = 60$) the degree distribution also displays a local maximum for highly connected nodes. This, however, might again be a finite-size effect as it disappears for a network with $n = 200$ (see Fig. 13). The simulations with networks of size $n = 200$ also generate much higher average and maximum degrees and result in more skewed degree distributions than in the case of $n = 60$. These variations of the degree distribution closely affect other characteristics of networks, such as the degree variance or the epidemic threshold of the network [2].

We find that the generated degree distributions $\bar{p}(k)$ mimic those in the real world strikingly well for suitable values of the balance threshold [22]. For example, compare the degree distribution for $\beta = 0.1$ and for $\beta = 0.4$ in Figure 13 with respectively the number of collaborators of movie actors and interlocking directorships (both reported in [22]).

## 5 Conclusions

The network model described above is based upon a plausible sociological concept — balance theory — and reproduces several characteristics of known social networks, notably a positive degree correlation and a variety of degree distributions. The model shows that networks evolve over time, and that their characteristic features require different periods of time until they reach their medium-term

stationary values. This, and the existence of different balance thresholds between groups, organisations, and populations might explain the observed variety of real world degree correlations and degree distributions. The reasonable fit with some empirical network properties can also be interpreted as a validation (but certainly not a proof) of balance theory. This line of reasoning offers several new avenues for empirical research. For example, it would be interesting to determine the average balance threshold in different contexts (nations, cultures, metropolitan vs. rural areas) and then verify the respective network properties like the degree correlation.

The model could be extended in many ways. For instance, it would be interesting to check out other values of $\alpha$ in order to test variations of *Structural Balance Theory* or to incorporate exogenous factors that influence the quality of links. Another extension could be to assign a distribution of balance thresholds across the nodes. Moreover, alternative network traits could be investigated and checked for their realism, especially on much larger networks. Finally, the model suggests that people constantly redefine their social contacts. Thus, the rate of social adjustments within the network (rebalancing sentiments by removing social links) is much faster than the rate at which new contacts are established. This contrasts with many other models of network evolution where new edges are added to the network at quicker rates than edges are removed (the usual logic being that links stay in the network for a person's lifetime) [5]. Interestingly, the gravity of social confrontations and revolutions (as, for example, measured by the number of workers involved in strikes [23] or by the number of victims in terrorists attacks [24]) seem to follow a power-law as well. If we therefore interpret the size of social upheavals as the change in the number of triangles of the underlying network, we can use the model as a conceptual bridge between the population's sentiments (tolerance), the evolution of its social network, and the likelihood of social disruptions.

We thank G. Brightwell, J. Howard, P. Sozou and two anonymous referees for helpful comments.

## References


1. M.E.J. Newman, SIAM Rev. **45**, 167 (2003)
2. R. Albert, A.-L. Barabási, Rev. Modern Phys. **74**, 47 (2002)
3. P. Doreian, Social Networks **24**, 93 (2002)
4. P. Doreian, D. Krackhardt, J. Math. Soc. **21**, 113 (2001)
5. H. Ebel, J. Davidsen, S. Bornholdt, Complexity **8**, 24 (2003)
6. F. Heider, J. Psychology **45**, 107 (1946)
7. D.C. Cartwright, F. Harary, Psychological Rev. **63**, 277 (1956)
8. S.F. Sampson, *A Novitiate in a Period of Change: An Experimental and Case Study of Social Relationships* (Cornell University, Ithaca, NY, 1968)
9. P. Doreian, A. Mrvar, Social Networks **18**, 149 (1996)
10. P. Bak, C. Tang, K. Wiesenfeld, Phys. Rev. A **38**, 364 (1987)
11. K.-I. Goh, D.-S. Lee, B. Kahng et al., Phys. Rev. Lett. **91**, 148701 (2003)
12. D. Hughes, M. Paczuski, R.O. Dendy et al., Phys. Rev. Lett. **90**, 131101 (2003)
13. G. Bianconi, G. Caldarelli, A. Capocci, Phys. Rev. E **71**, 066116 (2005)
14. K. Sneppen, M. Rosvall, A. Trusina et al., Europhyis. Lett. **67**, 349 (2004)
15. P. Fronczak, A. Fronczak, J.A. Holyst, Phys. Rev. E **73**, 046177 (2006)
16. Grinstein A, in *Scale Invariance, Interfaces, and Non-Equilibrium Dynamics*, edited by A. McKane et al. (Nato ASI Series, N.Y., 1995)
17. D. Sornette, in *Encyclopedia of Nonlinear Sciences*, edited by A. Scott (Routledge, N.Y., 2004)
18. Self-referring links are excluded
19. B. Bollobás, *Random Graphs* (Cambridge University Press, Cambridge, 2001)
20. M.E.J. Newman, Phys. Rev. Lett. **89**, 208701 (2002)
21. To substantiate this observation, however, one needs much larger networks and more data points for the distributions' tails
22. M.E.J. Newman, S.H. Strogatz, D.J. Watts, Phys. Rev. E **64**, 026118 (2001)
23. M. Biggs, Am. J. Soc. **110**, 1684 (2005)
24. A. Clauset, M. Young, K.S. Gleditsch, J. Conflict Resolution **51**, 1, 58 (2007)