Emergence of Good Conduct, Scaling and Zipf Laws in Human Behavioral Sequences in an Online World

Stefan Thurner\textsuperscript{1,2,3}, Michael Szell\textsuperscript{1}, Roberta Sinatra\textsuperscript{4,5}

\textsuperscript{1} Section for Science of Complex Systems, Medical University of Vienna, Vienna, Austria, \textsuperscript{2} Santa Fe Institute, Santa Fe, New Mexico, United States of America, \textsuperscript{3} IIASA, Laxenburg, Austria, \textsuperscript{4} Dipartimento di Fisica e Astronomia, Università di Catania and INFN, Catania, Italy, \textsuperscript{5} Laboratory for Complex Systems, Scuola Superiore di Catania, Catania, Italy

Abstract

We study behavioral action sequences of players in a massive multiplayer online game. In their virtual life players use eight basic actions which allow them to interact with each other. These actions are communication, trade, establishing or breaking friendships and enmities, attack, and punishment. We measure the probabilities for these actions conditional on previous taken and received actions and find a dramatic increase of negative behavior immediately after receiving negative actions. Similarly, positive behavior is intensified by receiving positive actions. We observe a tendency towards anti-persistence in communication sequences. Classifying actions as positive (good) and negative (bad) allows us to define binary ‘world lines’ of lives of individuals. Positive and negative actions are persistent and occur in clusters, indicated by large scaling exponents $\alpha \approx 0.87$ of the mean square displacement of the world lines. For all eight action types we find strong signs for high levels of repetitiveness, especially for negative actions. We partition behavioral sequences into segments of length $n$ (behavioral ‘words’ and ‘motifs’) and study their statistical properties. We find two approximate power laws in the word ranking distribution, one with an exponent of $\kappa \sim 1$ for the ranks up to 100, and another with a lower exponent for higher ranks. The Shannon $n$-tuple redundancy yields large values and increases in terms of word length, further underscoring the non-trivial statistical properties of behavioral sequences. On the collective, societal level the timeseries of particular actions per day can be understood by a simple mean-reverting log-normal model.

Introduction

Societies can be seen as individuals interacting through a multiplex network (MPN), i.e. a superposition of several social networks defined on the same set of nodes (individuals) [1,2]. Different types of networks correspond to different types of social interactions. For example the communication sub-network of the MPN is the network whose links correspond to the exchange of information by means of emails, telephone calls, or letters. Another subnetwork is the trading network, where goods or services are exchanged between individuals, in exchange for other goods, money, or –rarely– for nothing. Each of these interactions usually needs an initial action taken by one of the subjects involved in the exchange, the sender, and a target to receive it, the recipient. Actions can (but do not have to) be reciprocated, so that in general the MPN consists of a set of directed and weighted subnetworks. The MPN is a highly non-trivial dynamical object. The different social networks within the MPN are not independent but strongly influence each other through a network-network interaction. To understand systemic properties of societies it is essential to detect and quantify the organizational principles behind such mutual influences. The MPN is an example of a co-evolving structure: on one hand the actions of individuals shape and define the topological structure of the MPN. On the other hand the topology of the MPN constrains and influences the possible actions which take place on the MPN. In general the MPN of a society can not be observed due to immense requirements on synchronized data acquisition. Despite these difficulties, the analysis of small-scale MPNs has a tradition in the social sciences [1,3–5]. Concerning large-scale studies, recently there have been significant achievements in understanding a number of massive social networks on a quantitative basis, such as the cell phone communication network [6–8], features of the world-trade network [9,10], email networks [11], the network of financial debt [12] and the network of financial flows [13]. The integration of various dynamical networks of an entire society has so-far been beyond the scope of any realistic data source. However with the increasing availability of vast amounts of electronic fingerprints people leave throughout their lifes, this situation is about to change. Online sources are capturing more and more aspects of life, boosting our understanding of collective human behavior [14,15]. One particular source where complete behavioral multiplex data is available on the society level are massive multiplayer online games (MMOGs). In MMOGs hundreds of thousands of players meet online in a ‘virtual life’ where their actions can be easily studied [16]. Players have to gain their living through economic activity and usually are integrated in several types of social networks. In such games
Human behavioral sequences

We consider eight different actions every player can execute at any time. These are communication (C), trade (T), setting a friendship link (F), removing an enemy link (forgiving) (X), attack (A), placing a bounty on another player (punishment) (B), removing a friendship link (D), and setting an enemy link (E). While C, T, F and X can be associated with positive (good) actions, A, B, D and E are hostile or negative (bad) actions. We classify communication as positive because only a negligible part of communication takes place between enemies [18]. Segments of action sequences of three players (146, 199 and 701) are shown in the first three lines of Fig. 1 (a).

We consider three types of sequences for any particular player. The first is the stream of $N$ consecutive actions $A_i^i = \{a_i|n = 1, \ldots, N\}$ which player $i$ performs during his ‘life’ in the game. The second sequence is the (time-ordered) stream of actions that player $i$ receives from all the other players in the game, i.e., all the actions which are directed towards player $i$. We denote by $R_i = \{r_i|n = 1, \ldots, M\}$ received-action sequences. Finally, the third sequence is the time-ordered combination of player $i$’s actions and received-actions, which is a chronological sequence from the elements of $A_i^i$ and $R_i$ in the order of occurrence. The combined sequence we denote by $C_i^i$; its length is $M + N$, see also Fig. 1 (a). The $m$th element of one of these series is denoted by $A_i^i(n)$, $R_i(n)$, or $C_i^i(n)$. We do not consider the actual time between two consecutive actions which can range from milliseconds to weeks, rather we work in action-time.

If we assign $+1$ to any positive action C, T, F or X, and $−1$ to the negative actions A, B, D and E, we can translate a sequence $A_i^i$ into a symbolic binary sequence $A_i^\text{bin}$. From the cumulative sum of this sequence a ‘world line’ or ‘random walk’ for player $i$ can be generated, $W_\text{good-bad}(t) = \sum_{n=1}^{t} A_i^i(n)$, see Fig. 1 (b). Similarly, we define a binary sequence from the combined sequence $C_i^i$, where we assign $+1$ to an executed action and $−1$ to a received-action. This sequence we call $C_i^\text{bin}$, its cumulative sum, $W_\text{action}(t) = \sum_{n=1}^{t} C_i^\text{bin}(n)$ is the ‘action-receive’ random-walk or world line. Finally, we denote the number of actions which

\[
\text{(a) Actions}
\]

| Player 146 | ... A A A A A C C E E T E ... |
| Player 199 | ... C A C A A C F C C E A C ... |
| Player 701 | ... C C C C C C T T C C T C ... |

\[
\text{(b) Actions and received actions}
\]

| Player 199 | ... C A A C A A T F C T C C E E A C ... |

\[
\text{Figure 1. Short segment of action sequences of three players, A}^{146}, A^{199}, \text{and } A^{701}. (a) Some actions of players 146 and 701 are directed toward player 199. This results in a sequence of received-actions for player 199, R^{199} = \ldots \text{ATTCE} \ldots \text{.} \] The combined sequence of actions (originated from - and directed to) player 199, C^{199}, is shown in the last line: red letters mark actions from others directed to player 199. (b) Schematic illustration showing the definition of a binary walk in ‘good-bad’ action space (good-bad ‘world line’). A positive action (C, T, F or X) means an upward move, a negative action (A, B, D and E) is a downward move. Good people have rising world-lines.

\text{doi:10.1371/journal.pone.0029796.g001}
emergence of good conduct in human behavior

Table 1. First row: total number of actions by all players (with at least 1000 actions) in the Artemis universe of the Pardus game.

|       | D     | B     | A     | E     | F     | C     | T     | X     |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $N_{Y(d)}$ |       |       |       |       |       |       |       |       |
| mean  | 0.002 | 0.001 | 0.004 | -0.002 | -0.002 | 0.000 | 0.003 | 0.002 |
| std   | 1.13  | 0.79  | 0.54  | 0.64  | 0.35  | 0.12  | 0.28  | 0.94  |
| skew  | 0.12  | 0.26  | 0.35  | 0.08  | 0.23  | 0.11  | 1.00  | -0.01 |
| kurtosis | 3.35  | 3.84  | 6.23  | 3.67  | 3.41  | 3.76  | 13.89 | 3.72  |

Further rows: first 4 moments of $r_Y(d)$, the distribution of the log-increments of the $N_Y$ processes (see text). Approximate log-normality is indicated. The large values of kurtosis for $T$ and $A$ result from a few extreme outliers.

doi:10.1371/journal.pone.0029796.t001

occurred during a day in the game by $N_Y(d)$, where $d$ indicates the day and $Y$ stands for one of the eight actions.

Results

The number of occurrences of the various actions of all players over the entire time period is summarized in Tab. 1 (first line). Communication is the most dominant action, followed by attacks and trading which are each about an order of magnitude less frequent. The daily number of all communications, trades and attacks, $N_C(d)$, $N_T(d)$ and $N_A(d)$ is shown in Fig. 2 (a), (b) and (c), respectively. These processes are reverting around a mean, $R_Y$. All processes of actions show an approximate Gaussian statistic of its log-increments, $r_Y(d) = \log \frac{N_Y(d)}{N_Y(d-1)}$. The first 4 moments of the $r_Y$ series are listed in Tab. 1. The relatively large kurtosis for $T$ and $A$ results from a few extreme outliers. The distribution of log-increments for the $N_C$, $N_T$ and $N_A$ timeseries are shown in Fig. 2 (d). The lines are Gaussians for the respective mean and standard deviation from Tab. 1. As maybe the simplest mean-reverting model with approximate log-normal distributions, we propose

$$N_Y(d) = N_Y(d-1) e^{(d-1)R_Y} e^{(d-1)\rho_Y}$$

(1)

where $\rho_Y$ is the mean reversion coefficient, $\xi(d)$ is a realization of a zero mean Gaussian random number with standard deviation $\sigma_Y$, and $R_Y$ is the value to which the process $N_Y(t)$ reverts to. $\sigma$ is given by the third line in Tab. 1. Note that this is an AR(1) process (Ornstein-Uhlenbeck process in discrete time) in logarithmic variables.

Transition probabilities

With $p(Y|Z)$ we denote the probability that an action of type $Y$ follows an action of type $Z$ in the behavioral sequence of a player. $Y$ and $Z$ stand for any of the eight actions, executed or received (received is indicated by a subscript $r$). In Fig. 3 (a) the transition probability matrix $p(Y|Z)$ is shown. The $y$ axis of the matrix indicates the action (or received-action) happening at a time $t$, the probabilities for the actions (or received-actions) that immediately follow are given in the corresponding horizontal place.

This transition matrix specifies to which extent an action or a received action of a player is influenced by the action that was done or received at the previous time-step. In fact, if the behavioral sequences of players had no correlations, i.e. the probability of an action, received or executed, is independent of the history of the player’s actions, the transition probability $p(Y|Z)$ simply is $p(Y)$, i.e. to the probability that an action or received action $Y$ occurs in the sequence is determined by its relative frequency only. Therefore, deviations of the ratio $\frac{p(Y|Z)}{p(Y)}$
from 1 indicate correlations in sequences. In Fig. 3 (b) we report the values of \( p_r(Y|Z) \) for actions and received actions (received actions are indicated with the subscript \( r \)) classified only according to their positive (+) or negative (−) connotation. In brackets we report the \( \zeta \)-score with respect to the uncorrelated case. We find that the probability to perform a good action is significantly higher if at the previous time-step a positive action has been received. Similarly, it is more likely that a player is the target of a positive action if at the previous time-step he executed a positive action. Conversely, it is highly unlikely that after a good action, executed or received, a player acts negatively or is the target of a negative action. Instead, in the case a player acts negatively, it is most likely that he will perform another negative action at the following time-step, while it is highly improbable that the following action, executed or received, will be positive. Finally, in the case a negative action is received, it is likely that another negative action will be received at the following time-step, while all other possible actions and received actions are under-represented. The high statistical significance of the cases \( P_r(−|−) \) hints toward a high persistence of negative actions in the players' behavior, see below.

Another finding is obtained by considering only pairs of received actions followed by performed actions. This approach allows to quantify the influence of received actions on the performed actions of players. For these pairs we measure a probability of 0.02 of performing a negative action after a received positive action. This value is significantly lower compared to the probability of 0.10 obtained for randomly reshuffled sequences. Similarly, we measure a probability of 0.27 of performing a negative action after a received negative action. Note that this result is not in contrast with the values in Fig. 3 (b), since only pairs made up of received actions and performed actions are taken into account. Our results agree with a recent study where the emotional content of posts in online forums was analyzed similarly [24].

### World lines

The world lines \( W_{good-bad} \) of good-bad action sequences are shown in Fig. 4 (a), the action-reaction world lines in Fig. 4 (b). As a simple measure to characterize these world lines we define the slope \( k \) of the line connecting the origin of the world line to its end point (last action of the player). A slope of \( k = 1(−1) \) in the good-bad world lines \( W_{good-bad} \) indicates that the player performed only positive (negative) actions. The slope \( k' \) is an approximate measure of ‘altruism’ for player \( i \). The histogram of the slopes for all players is shown in Fig. 4 (b), separated into good (blue) and bad (red) players, i.e., players who have performed more good than bad actions and vice versa. The mean and standard deviation of slopes of good, bad, and all players are \( k_{good} = 0.81 \pm 0.19 \), \( k_{bad} = −0.40 \pm 0.28 \), and \( k_{all} = 0.76 \pm 0.31 \), respectively. Simulated random walks with the same probability 0.90 of performing a positive action yield a much lower variation, \( k_{sim} = 0.81 \pm 0.01 \), pointing at an inherent heterogeneity of human behavior. For the combined action-received-action world line \( W_{act-rec} \) the slope is a measure of how well a person is integrated in her social environment. If \( k = 1 \) the person only acts and receives no input, she is ‘isolated’ but dominant. If the slope is \( k = −1 \) the person is driven by the actions of others and does never act nor react. The histogram of slopes for all players is shown in Fig. 4 (c). Most players are well within the ±45 degree cone. Mean and standard deviation of slopes of good, bad, and all players are \( k_{good} = 0.02 \pm 0.10 \), \( k_{bad} = 0.30 \pm 0.19 \), and \( k_{all} = 0.04 \pm 0.12 \), respectively. Bad players are tendentially dominant, i.e., they perform significantly more actions than they receive. Simulated random walks with equal probabilities for up and down moves for a sample of the same sequence lengths, we find again a much narrower distribution with slope \( k_{sim} = 0.00 \pm 0.01 \).

As a second measure we use the mean square displacement of world lines to quantify the persistence of action sequences,

\[
M^2(t) = \langle \Delta W(t) - \langle \Delta W(t) \rangle \rangle^2 \propto t^2, \tag{2}
\]

\[
\Delta W(t) = \sum_{t=1}^{T} [W(t+1) - W(t)]
\]
for the good-bad world lines are slight tendency towards persistence. Mean and standard deviation strongly persistent behavior is obvious, in the second there is a separated into good (blue) and bad (red) players, is shown in Fig. 4 world lines that this fit range is indeed reasonable. 100. We checked from the mean square displacement of single calculating the exponents above (below) 0 in (a) and only fall into the \( k > 0 \) (\( k < 0 \)) category in (b). (d) World lines of action-received random walks, (e) distribution of their slopes \( k \) and (f) of their scaling exponents \( \alpha \). The inset in (d) shows only the world lines of bad players. These players are typically dominant, i.e. they perform significantly more actions than they receive. In total the players perform many more good than bad actions and are strongly persistent with good as well as with bad behavior, see (c), i.e. actions of the same type are likely to be repeated.

\[ \Delta W(t+\tau) - W(t) \approx k \]

where \( \Delta W(t+\tau) - W(t) \) is the average over all \( t \). The asymptotic behavior of \( M(t) \) yields information about the ‘persistence’ of a world line. \( M(t) \propto t^z \) is the pure diffusion case, \( M(t) \propto t^z \) with scaling exponent \( z \neq \frac{1}{2} \) indicates persistence for \( z > \frac{1}{2} \), and anti-persistence for \( z < \frac{1}{2} \). Persistence means that the probability of making an up(down) move at time \( t+1 \) is larger(less) than \( p = 1/2 \), if the move at time \( t \) was an up move. For calculating the exponents \( z \) we use a fit range of \( t \) between 5 and 100. We checked from the mean square displacement of single world lines that this fit range is indeed reasonable.

The histogram of exponents \( z \) for the good-bad random walk, separated into good (blue) and bad (red) players, is shown in Fig. 4 (c), for the action-received-action world line in (l). In the first case strongly persistent behavior is obvious, in the second there is a slight tendency towards persistence. Mean and standard deviation for the good-bad world lines are \( z_{\text{good-bad}} = 0.87 \pm 0.06 \), for the action-received actions \( z_{\text{act-rec}} = 0.59 \pm 0.10 \). Simulated sequences of random walks have – as expected by definition – an exponent of \( z_{\text{rand}} = 0.5 \), again with a very small standard deviation of about 0.02. Figure 4 (a) also indicates that the lifetime of players who use negative actions frequently is short. The average lifetime for players with a slope \( k < 0 \) is \( 2528 \pm 1856 \) actions, compared to players with a slope \( k > 0 \) with \( 3909 \pm 4559 \) actions. The average lifetime of the whole sample of players is \( 3849 \pm 4484 \) actions.

Motifs, Entropy and Zipf law

By considering all the sequences of actions \( A^i \) of all possible players \( i \), we have an ensemble which allows to perform a motif analysis [25]. We define a \( n \)-string as a subsequence of \( n \) contiguous actions. An \( n \)-motif is an \( n \)-string which appears in the sequences with a probability higher than expected, after lower-order correlations have been properly removed.

Across the entire ensemble, \( 8^8 \) different \( n \)-strings can appear, each of them occurring with a different probability. The frequency, or observed probability, of each \( n \)-string can be compared to its expected probability of occurrence, which can be estimated on the basis of the observed probability of lower order strings, i.e. on the frequency of \((n-1)\)-strings. For example, the expected probability of occurrence of a 2-string \((A_i,A_{i+1})\) is estimated as the product of the observed probability of the single actions \( A_i \) and \( A_{i+1} \), \( p^{\text{obs}}(A_i,A_{i+1}) = p^{\text{obs}}(A_i)p^{\text{obs}}(A_{i+1}) \). Similarly, the probability of a 3-string \((A_i,A_{i+1},A_{i+2})\) to occur can be estimated as \( p^{\text{exp}}(A_i,A_{i+1},A_{i+2}) = p^{\text{obs}}(A_i,A_{i+1})p^{\text{obs}}(A_{i+2}|A_{i+1}) \), where \( p^{\text{obs}}(A_{i+2}|A_{i+1}) \) is the conditional probability to have action \( A_{i+2} \) following action \( A_{i+1} \). By definition of conditional probability, one has \( p^{\text{obs}}(A_{i+2}|A_{i+1}) = p^{\text{obs}}(A_{i+1},A_{i+2})/p^{\text{obs}}(A_{i+1}) \) (see [25] for details). A \( n \)-motif in the ensemble is then defined as a \( n \)-string whose observed probability of occurrence is significantly higher than its expected probability.

We computed the expected and observed probabilities \( p^{\text{obs}} \) and \( p^{\text{exp}} \) for all \( 8^2 = 64 \) 2-strings and for all \( 8^3 = 512 \) 3-strings, focusing on those \( n \)-strings with the highest ratio \( \frac{p^{\text{obs}}}{p^{\text{exp}}} \). Higher orders are statistically not feasible due to combinatorial explosion. We find that the 2-motifs in the sequences of actions \( A \) are clusters of same letters: BB, DD, XX, EE, FF, AA with ratios \( \frac{p^{\text{obs}}}{p^{\text{exp}}} \approx 169, 136, 117, 31, 15, 10 \), respectively. This observation is consistent with the previous first-order observation that actions cluster. The most significant 3-motifs however are (with two exceptions) palin-
The analysis of human behavioral sequences as recorded in a massive multiplayer online game shows that communication is by far the most dominant activity followed by aggression and trade. Communication events are about an order of magnitude more frequent than attacks and trading events, showing the importance of information exchange between humans. It is possible to understand the collective timeseries of human actions of a particular type ($N_T$) with a simple mean-reverting log-normal model. On the individual level we are able to identify organizational patterns of the emergence of good overall behavior. Transition rates of actions of individuals show that positive actions strongly induces positive *reactions*. Negative behavior on the other hand has a high tendency of being repeated instead of being reciprocated, showing the ‘propulsive’ nature of negative actions. However, if we consider only reactions to negative actions, we find that negative reactions are highly over-represented. The probability of acting out negative actions is about 10 times higher if a person received a negative action at the previous timestep than if she received a positive action. The action of communication is found to be of highly reciprocal ‘back-and-forth’ nature. The analysis of binary timeseries of players (good-bad) shows that the behavior of almost all players is ‘good’ almost all the time. Negative actions are balanced to a large extent by good ones. Players with a high fraction of negative actions tend to have a significantly shorter life. This may be due to two reasons: First because they are hunted down by others and give up playing, second because they are unable to maintain a social life and quit the game because of loneliness or frustration. We interpret these findings as empirical evidence for self organization towards reciprocal, good conduct within a human society. Note that the game allows bad behavior in the same way as good behavior but the extent of punishment of bad behavior is freely decided by the players.

Behavior is highly persistent in terms of good and bad, as seen in the scaling exponent ($\sim 0.87$) of the mean square displacement of the good-bad world lines. This high persistence means that good and bad actions are carried out in clusters. Similarly high levels of persistence were found in a recent study of human behavior [28]. A smaller exponent ($\sim 0.59$) is found for the action–received-action timeseries.

Finally we split behavioral sequences of individuals into subsequences (of length 1–6) and interpret these as behavioral ‘words’. In the ranking distribution of these words we find a Zipf law of the form $P_r \sim r^{-\kappa}$. The Shannon word redundancy [29] (see e.g. [20–22]) for symbol sequences composed of 8 symbols (our action types) is defined as

$$R(n) = 1 - \frac{1}{\text{log}_2 n!} \sum_i n^{-1} \log (p_i^{\text{obs}}),$$

where $p_i^{\text{obs}}$ is the probability of finding a specific $n$-letter word. Uncorrelated sequences yield an equi-distribution, $p_i = 8^{-n}$, i.e. $R(0) = 0$. In the other extreme of only one letter being used, $R(0) = 1$. In (3) $R(n)$ is shown as a function of sequence length $n$. Shannon redundancy is not a constant but increases with $n$. This indicates that Boltzmann-Gibbs entropy might not be an extensive quantity for action sequences [27].

### Discussion

The analysis of human behavioral sequences as recorded in a massive multiplayer online game shows that communication is by far the most dominant activity followed by aggression and trade. Communication events are about an order of magnitude more frequent than attacks and trading events, showing the importance of information exchange between humans. It is possible to understand the collective timeseries of human actions of a particular type ($N_T$) with a simple mean-reverting log-normal model. On the individual level we are able to identify organizational patterns of the emergence of good overall behavior. Transition rates of actions of individuals show that positive actions strongly induces positive *reactions*. Negative behavior on the other hand has a high tendency of being repeated instead of being reciprocated, showing the ‘propulsive’ nature of negative actions. However, if we consider only reactions to negative actions, we find that negative reactions are highly over-represented. The probability of acting out negative actions is about 10 times higher if a person received a negative action at the previous timestep than if she received a positive action. The action of communication is found to be of highly reciprocal ‘back-and-forth’ nature. The analysis of binary timeseries of players (good-bad) shows that the behavior of almost all players is ‘good’ almost all the time. Negative actions are balanced to a large extent by good ones. Players with a high fraction of negative actions tend to have a significantly shorter life. This may be due to two reasons: First because they are hunted down by others and give up playing, second because they are unable to maintain a social life and quit the game because of loneliness or frustration. We interpret these findings as empirical evidence for self organization towards reciprocal, good conduct within a human society. Note that the game allows bad behavior in the same way as good behavior but the extent of punishment of bad behavior is freely decided by the players. Behavior is highly persistent in terms of good and bad, as seen in the scaling exponent ($\sim 0.87$) of the mean square displacement of the good-bad world lines. This high persistence means that good and bad actions are carried out in clusters. Similarly high levels of persistence were found in a recent study of human behavior [28]. A smaller exponent ($\sim 0.59$) is found for the action–received-action timeseries.

Finally we split behavioral sequences of individuals into subsequences (of length 1–6) and interpret these as behavioral ‘words’. In the ranking distribution of these words we find a Zipf law of the form $P_r \sim r^{-\kappa}$. The Shannon word redundancy [29] (see e.g. [20–22]) for symbol sequences composed of 8 symbols (our action types) is defined as

$$R(n) = 1 - \frac{1}{\text{log}_2 n!} \sum_i n^{-1} \log (p_i^{\text{obs}}).$$

where $p_i^{\text{obs}}$ is the probability of finding a specific $n$-letter word. Uncorrelated sequences yield an equi-distribution, $p_i = 8^{-n}$, i.e. $R(0) = 0$. In the other extreme of only one letter being used, $R(0) = 1$. In (3) $R(n)$ is shown as a function of sequence length $n$. Shannon redundancy is not a constant but increases with $n$. This indicates that Boltzmann-Gibbs entropy might not be an extensive quantity for action sequences [27].

### Acknowledgments

We thank Werner Bayer for compiling Pardus data.

### Author Contributions

Conceived and designed the experiments: ST MS RS. Performed the experiments: ST MS RS. Analyzed the data: ST MS RS. Contributed reagents/materials/analysis tools: ST MS RS. Wrote the paper: ST.

### References

1. Wasserman S, Faust K (1994) Social Network Analysis: Methods and Applications Cambridge Univ Press.
2. Szell M, Lambiotte R, Thurner S (2010) Multirelational organization of large-scale social networks in an online world. Proc Natl Acad Sci USA 107: 13630–13641.
3. McPherson M, Smith-Lovin L, Cook J (2001) Birds of a feather: Homophily in social networks. Annu Rev of Sociol. pp 415–444.

4. Entwisle B, Faust K, Rindfuss R, Kaneda T (2007) Networks and contexts: Variation in the structure of social ties. Am J Sociol 112: 1495–1533.

5. Padgett J, Ansell C (1993) Robust action and the rise of the Medici, 1400–1434. Am J Sociol. pp 1259–1319.

6. Onnela J, Saramäki J, Hyvönen J, Szabó G, de Menezes M, et al. (2007) Analysis of a large-scale weighted network of one-to-one human communication. New J of Physics 9: 179.

7. Onnela J, Saramäki J, Hyvönen J, Szabó G, Lazer D, et al. (2007) Structure and tie strengths in mobile communication networks. Proc Natl Acad Sci USA 104: 7332.

8. Lambiotte R, Blondel V, de Kerchove C, Huser E, Prieur C, et al. (2008) Geographical dispersal of mobile communication networks. Physica A 387: 5317–5325.

9. Hidalgo C, Klinger B, Barabási AL, Hausmann R (2007) The product space conditions the development of nations. Science 317: 482.

10. Hidalgo C, Hausmann R (2009) The building blocks of economic complexity. Proc Natl Acad Sci USA 106: 10570–10575.

11. Newman M, Forrest S, Balthrop J (2002) Email networks and the spread of computer viruses. Phys Rev E 66: 035101.

12. Boss M, Elsinger H, Summer M, Thurner S (2004) Network topology of the interbank market. Quantitative Finance 4: 677–684.

13. Kyriakopoulos F, Thurner S, Puhar C, Schmutz S (2009) Network and eigenvalue analysis of financial transaction networks. Eur Phys J B 71: 523–531.

14. Lazer D, Pentland A, Adamic L, Aral S, Barabási AL, et al. (2009) Computational social science. Science 323: 721.

15. Lewis K, Kaufman J, Gonzalez M, Wimmer A, Christakis N (2008) Tastes, ties, and time: A new social network dataset using Facebook.com. Social Networks 30: 330–342.

16. Bainbridge W (2007) The scientific research potential of virtual worlds. Science 317: 472.

17. Grabowski A, Krużewska N (2007) Experimental study of the structure of a social network and human dynamics in a virtual society. Int J of Mod Phys C 18: 1527–1536.

18. Szell M, Thurner S (2010) Measuring social dynamics in a massive multiplayer online game. Social Networks 32: 313–329.

19. Pardus website. Available at http://www.pardus.at. Accessed 2011 Dec 12.

20. Stanley H, Buldyrev S, Goldberger A, Havlin S, Peng C, et al. (1999) Scaling features of noncoding DNA. Physica A 273: 1–18.

21. Peng G, Buldyrev S, Goldberger A, Havlin S, Sciortino F, et al. (1992) Long-range correlations in nucleotide sequences. Nature 356: 168–170.

22. Peng G, Buldyrev S, Havlin S, Simons M, Stanley H, et al. (1994) Mosaic organization of dna nucleotides. Phys Rev E 49: 1605.

23. Castronova E (2005) Synthetic Worlds: The Business and Culture of Online Games. Chicago: University of Chicago Press.

24. Chmiel A, Sienkiewicz J, Thelwall M, Paltoglou G, Buckley K, et al. (2011) Collective emotions online and their influence on community life. PloS one 6: e22207.

25. Sinatra R, Condorelli D, Latora V (2010) Networks of motifs from sequences of symbols. Phys Rev Lett 105: 178702.

26. Zipf G (1949) Human behavior and the principle of least effort. Addison-Wesley.

27. Hanel R, Thurner S (2011) A comprehensive classification of complex statistical systems and an axiomatic derivation of their entropy and distribution functions. Eur Phys Lett 93: 20006.

28. Fan C, Guo J, Zha Y (2010) Fractal analysis on human behaviors dynamics. arXiv. pp 10124000v1.