Human mobility in an online world

Working paper SFI summer school 2010, last update 2010-08-30

Michael Szell *,†, **Giovanni Petri** ‡ §, **Kang Zhao** ¶ **and Drew Levin** ‖

*Section for Science of Complex Systems, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria,‡Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London, London, SW7 2AZ, UK,§Institute for Mathematical Sciences, Imperial College London, London SW7 2PG, UK,¶College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA 16802, USA, and ‖Department of Computer Science, University of New Mexico, Albuquerque, NM 87131, USA

Massive multiplayer online games provide a fascinating new way of observing hundreds of thousands of simultaneously interacting individuals engaged in virtual socio-economic activities. We have compiled a data set consisting of practically all actions of all players over a period of three years from an online game played by over 350,000 people. The universe of this online world is a network on which players move to interact with other players. This interaction may consist of trade, armed conflict, friendship and enmity. We focus on the mobility of human players on the network over a time-period of 500 days. We take a number of mobility measurements (daily and biweekly position changes, entropy, number of unique nodes visited) of players and compare them with measures of simulated random walkers on the same topology. Player mobility is highly different from the mobility of unbiased random walkers. The analysis of biased random walkers reveals the two essential ingredients which explain measured human mobility patterns most accurately: heterogeneity and a tendency to return to recently visited locations. We compare our entropy distributions with human mobility in real life world – measured via mobile phone data – and find a striking match.

random walk | complex system | online game

Human society constitutes a complex system extending over several dimensions. First, humans stand in relation to a multitude of other humans, spanning a number of socio-economic networks of various relations and intensities, such as family, work, friendship or enmity. Second, individuals move in a spatial dimension, either by foot in euclidian space, by car on street networks, or on transport networks such as the network spanned by airports or underground stations. Third, a necessary dimension for all processes involving human behavior is time.

A prime example for a single process taking place on whole societies for which all of these three components have to be accounted for is the spread of epidemics. For epidemics to propagate, humans need to be close in space and time which is usually induced by social contacts. The role of the underlying social networks [1] is as important as the spatial topology on which humans move [2], as is the order of interaction in time [3]. Human mobility is obviously a closely linked and therefore important issue to understand. A number of behavioral and mobility studies from a complex systems perspective [4, 5, 6, 7, 8, 9, 10, 11] have shown a variety of rich phenomena in everyday human behavior and mobility. As has been pointed out, the observed patterns often display relatively high regularity and potential predictability [9]. While the majority of these studies have been conducted on mobile phone data, there are alternative and for some aspects richer data sets available.

The online game 'Pardus'. Massive multiplayer online games (MMOG) provide a fascinating new way of observing hundreds of thousands of simultaneously socially interacting individuals engaged in virtual economic activities [12, 13, 14, 15]. We have compiled a data set consisting of practically all actions of all players over a period of three years from a MMOG played by over 300,000 people. Here we focus on the movement of players in the game's virtual universe, to shed light on mobility patterns within human societies.

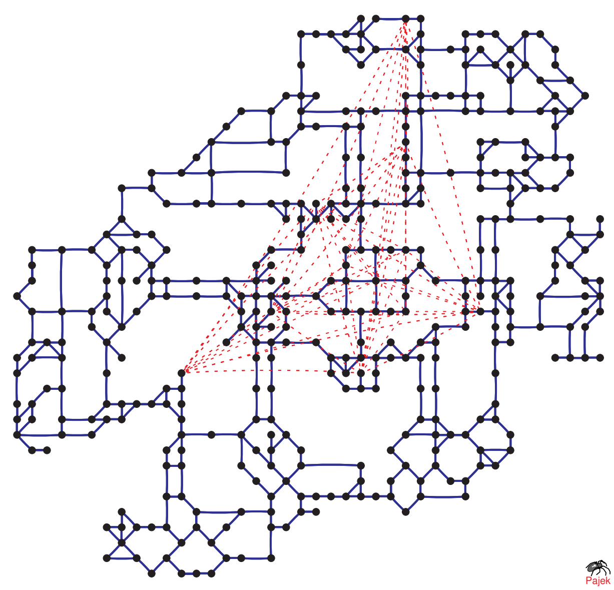

Fig. 1. Map of the Pardus universe which has a network topology. Nodes represent sectors. Wormholes connect nearby sectors (solid lines), seven X-holes connect far apart sectors (dashed lines). Players move between sectors to get involved and to interact with other players.

Pardus (www.pardus.at) is a browser-based MMOG in a science-fiction setting, open to the public since September 2004. A browser-based MMOG is characterized by a substantial number of users playing together in the same virtual environment connected by an internet browser. For a detailed categorization of online games see [16, 17].

In Pardus every player owns an account with one *character* per game universe. A character is a pilot owning a spacecraft with a certain cargo capacity, roaming the virtual universe trading commodities, socializing, and much more, 'to gain wealth and fame in space'. The main component of Pardus consists of trade with a society of players heavily driven by social factors such as friendship, cooperation or competition. There is no explicit 'winning' in Pardus as there is no inherent set of goals. Pardus is a *virtual world* whose gameplay is based on socializing and role-playing, with interaction

† To whom correspondence should be addressed. E-mail: michael.szell@meduniwien.ac.at

Table 1. Network properties of the Pardus universe: number of nodes N, number of edges L, average degree $\bar{k}$, clustering coefficient C, clustering coefficient to corresponding coefficient of random graph C/C_{r}, average geodesic $\bar{g}$, average geodesic to corresponding average geodesic of random graph $\bar{g}/\bar{g}_{\mathrm{r}}$, local efficiency E_{loc}, global efficiency E_{glob}, diameter D, effective diameter (0.9-quantile) D_{eff}.

N	400
L	1160
$\bar{k}$	2.9
C	0.089
C/C_{r}	12.33
$\bar{g}$	11.89
$\bar{g}/\bar{g}_{\mathrm{r}}$	2.11
E_{loc}	0.80
E_{glob}	0.03
D	27
D_{eff}	18

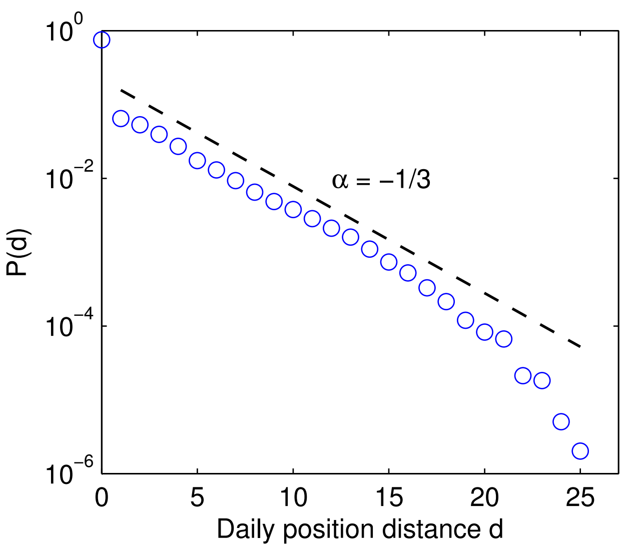

Fig. 2. Probability distribution of daily position distances, for all 1991 players and all 498 days. The dashed line is an exponential with slope $\alpha = -1/3$.

between players and non-player characters as its core elements [17].

A Pardus universe consists of 400 sectors, where each sector is a rectangular grid made up of an average of 15×15 fields. A field is the atomic unit of space between which players move. Each sector has a unique sector-id, each field has a unique field-id, and each field-id corresponds to a sector-id (because a sector consists of several fields, there are multiple field-ids corresponding to the same sector-id). To players, sectors are displayed as nodes on a two-dimensional cartesian coordinate system, the universe map, see Fig. 1. Sectors adjacent on the universe map may be interconnected by wormhole-links, allowing players to hop from sector to sector. Therefore, the universe is a network, where nodes are sectors and links are wormholes. While wormhole links always link together exactly two sectors (solid lines in Fig. 1), so-called X-holes connect to all other X-holes in the universe (dashed lines in Fig. 1). This results in a "folding" of space, with X-holes bringing far apart sectors closer together. Players incur a movement cost for traveling fields and for jumping through wormholes and X-holes, limiting their daily movement.

Results

Universe network. Network properties of the Pardus universe are given in Table . The universe is sparse ($\bar{k} = 2.9 \ll 400 = N$), highly clustered ($C/C_{\mathrm{r}} = 12.33$), but does not exhibit small-world features [18]: The average geodesic is not close to the random graph average geodesic ($\bar{g}/\bar{g}_{\mathrm{r}} = 2.11$) [19], the diameter is relatively big ($D = 27$), and the global efficiency is low ($E_{\mathrm{glob}} = 0.03$) [20]. Excluding X-hole links, the universe is a planar network, i.e. it has no intersecting links, thus it is similar to euclidian networks such as street traffic networks. Locally, the universe is lattice-like, which is reflected in the relatively high local efficiency ($E_{\mathrm{loc}} = 0.80$).

Player movement. Daily snapshots of player positions (see Materials and Methods) allow us to measure daily position distances of all players on the universe network. Due to lack of high-frequency data available, we treat these "jumps" as following shortest paths (which seems a reasonable assumption since players usually try to minimize movement costs). Figure 2 shows the probability distribution of daily position distances for all 1991 players, over the whole timespan of 498 days. On average, a player is stationary on $\approx 75\%$ of all days. When a player does move, the distribution representing the distance covered d decays roughly exponentially with slope $\alpha = -1/3$:

$$d \sim e^{-\frac{1}{3}d} \qquad [1]$$

Note that this distribution is taken over the movement activity of *all* players, thus a heterogeneity in player activity and mobility is not accounted for.

Figure 3 shows the relation between activity – measured in Action Points (APs) used (see Materials and Methods) – and mobility of all 1991 players. High activity is necessary, but not sufficient, for high mobility: in order to have a high mobility, a player needs to spend many APs, but there exists also a large number of players with low mobility who spend a high number of APs (in actions other than travelling).

Besides average daily changes in position, we calculated biweekly position changes (i.e. distance in position between day d and day $d + 14$ for all days), random and spatial entropy (see Materials and Methods). Probability distributions of these properties are shown in the first row of Fig. 5. Note the spike at zero for daily position distances, which correspond to 31 players whose measured positions have never changed over all 498 days, despite some of them having spent a considerable number of action points.

Comparison with random walkers. To gather insights into the complexity of the movement process of Pardus players, we compare their mobility patterns with those of different kinds of random walkers, on the same network.

Unbiased random walkers. We first start with the most simple kind of random walkers, namely unbiased random walkers (URW) which start on a random node and each timestep choose one of their neighbours at random with equal probability. As has been proven analytically [21], the stationary solution of URWs on networks is, up to normalization,

[1] In case that this is an empty set, the first non-empty set of sectors with largest distance smaller than d is selected.

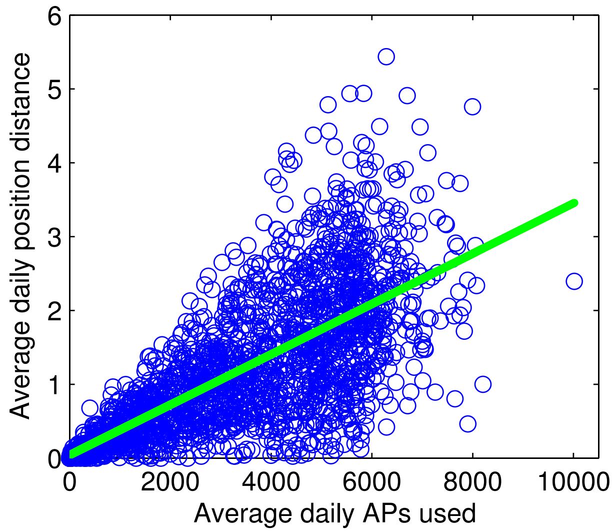

Fig. 3. Average daily position distances versus Action Points used (see Materials and Methods), for all 1991 players. The line is a least-squares fit, making a trend visible between mobility and activity. However, the distribution is very widespread: There are many highly active players with a low mobility.

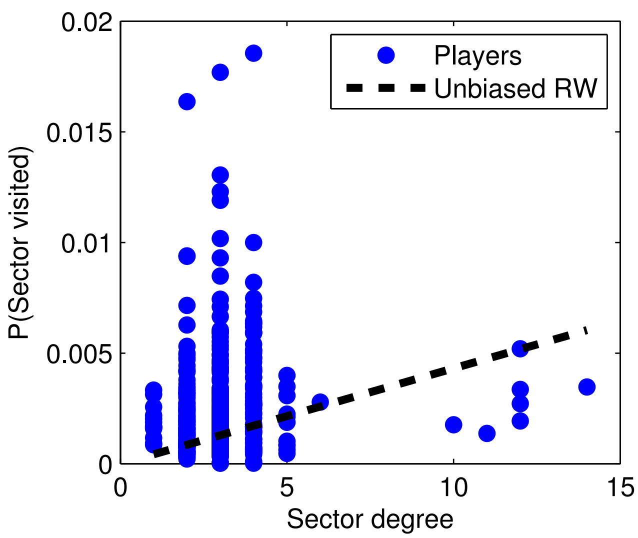

Fig. 4. Probability that a sector was visited versus sector degree for all 400 sectors, shown for players (discs) and unbiased random walkers (dashed line, analytical result). Players have a strong tendency to visit a few specific sectors, independent of degree. In this respect, mobility of players deviates strongly from unbiased random walkers.

identical to the degree of the nodes. For example, in the long run a node with degree $2k$ will be visited twice as often by URWs than a node with degree k. Comparison of the probability of visits between URWs and the movement behavior of players, Fig. 4, shows that in this respect players deviate strongly from the URW null model. In particular, the few existing high-degree sectors ($k \geq 10$) do not display a preference for being visited by players, while a few low-degree sectors are visited with a significantly higher preference than other similar-degree sectors. This can be explained by the game's setup, where visiting specific special sectors more often may be important to players for a variety of reasons (such as sectors containing special resources).

Biased homogeneous random walkers. Since there exists a wide range of daily position distances covered by players (including standstill), unbiased random walkers which hop exactly one node apart every fixed timestep are obviously not able to explain the observed human mobility patterns. For this reason we conduct random walks with a set of random walkers whose behavior fits adequately the sampling bias of daily-only data. We use 1991 random walkers, starting on the same positions as the 1991 players in the available data, and simulate for 498 timesteps, each one corresponding to a day. Every timestep, each random walker draws a distance d from the measured distribution of daily position distances, Fig. 2, and selects a random destination sector from all sectors which are d hops away from its current position[1]. We call these biased random walkers *homogeneous* because they all follow equivalent rules.

The second row in Fig. 5 shows probability distributions of properties of these biased homogeneous walkers. Compared to the distributions of player properties in the first row, the distributions of these random walkers are very narrow, giving evidence for a high heterogeneity in mobility behavior of players. Note how the maxima of all shown property distributions of the random walkers occur always at far higher values than the maxima in property distributions of players.

Biased heterogeneous random walkers. Because the biased homogeneous walkers exhibit very narrow distributions, we now add a heterogeneous component. Again we use 1991 random walkers with starting positions of players, over 498 timesteps. This time however, we give each random walker the distribution of daily position distances from a single corresponding player to draw from. Taken together, all random walkers again exhibit the same distribution of daily position distances, only in this case, each random walker is unique. Similar to the human players, 31 of the walkers stand still for all simulated 498 days, while some move rarely, and some move very frequently and over long distances.

As the third row in Fig. 5 shows, property distributions of heterogeneous random walkers are much broader than the homogeneous ones. By construction, daily position distances are identical to players (ignoring fluctuations), Fig. 5 (i). The heterogeneous walkers generate broader distributions (in fact they are too broad), but still cannot match the other properties of human players. Distribution maxima of heterogeneous walkers can be found at similar high values as homogeneous walkers. In practice, the deviation in biweekly position distances, Fig. 5 (j), means that players are *more likely to stay in the vicinity of recently visited sectors* than expected by chance. The deviation in random entropy, Fig. 5 (k), can – as in the case of unbiased random walkers – be explained by the highly specific importance assigned by players to staying in certain sectors. These effects lead to a smaller average spatial entropy for players than would be expected by chance.

Biased heterogeneous random walkers with return to previous sector. We implement a final refinement to the behavior of the biased random walkers, motivated by the following additional measurements. So far, we implicitly as-

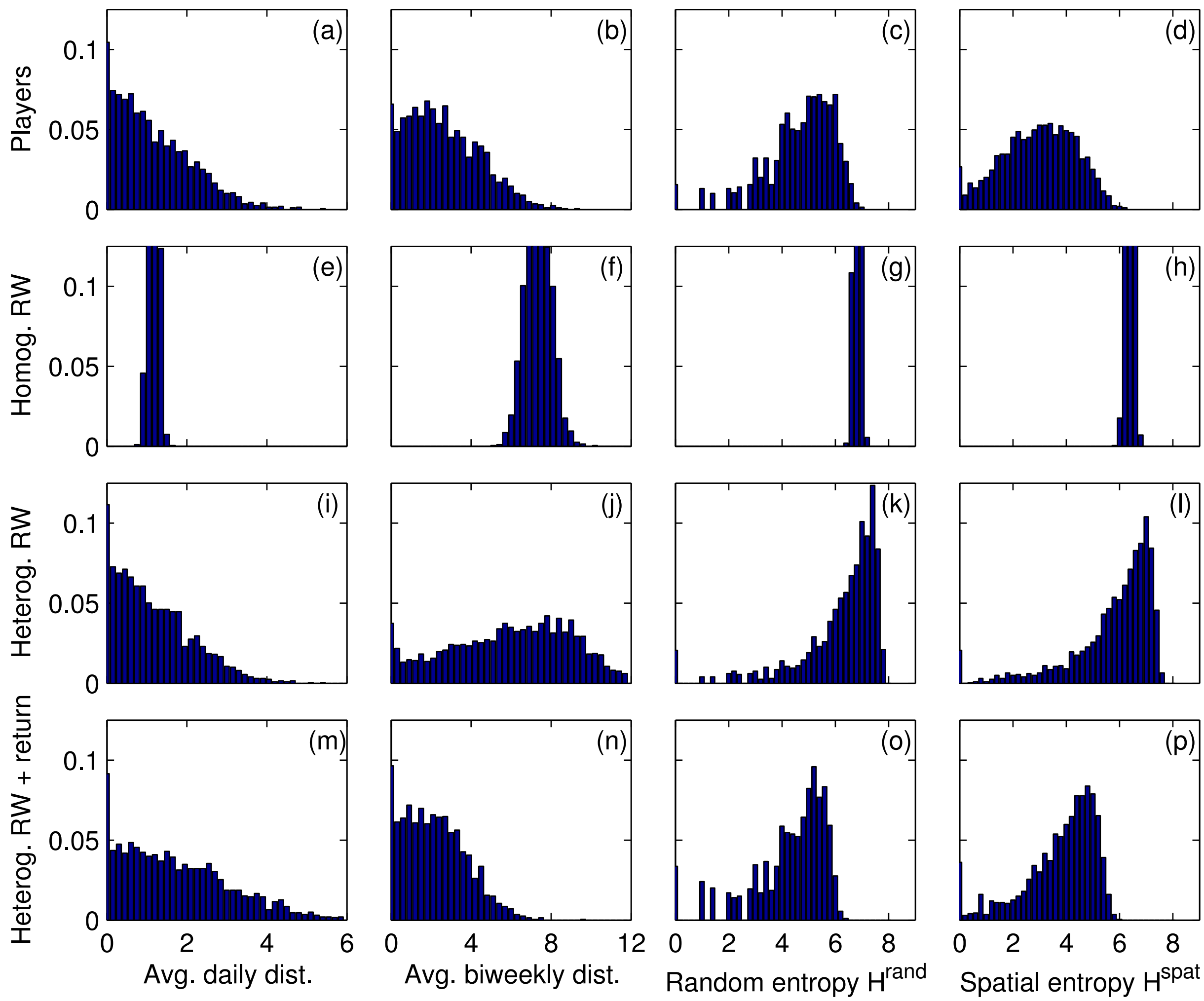

Fig. 5. Probability distributions of average daily (column 1) and biweekly (column 2) distances in positions, random entropy H^{rand} (column 3), and spatial entropy H^{spat} (column 4) for 1991 players (row 1), biased homogeneous random walkers (row 2) biased heterogeneous random walkers (row 3), and biased heterogeneous walkers who return to the previously visited sector 80% of the time.

sumed that a position change of a player does not depend on her past position changes and does not influence her future position changes. To shed light upon whether position changes are independent or not, we relate distances between subsequent position changes of players. For consecutive position changes – from a first position p_1 to a second one p_2, then (after a possible time of standstill) to a third position p_3 – we measured the shortest path distances $d(p_1, p_2)$, $d(p_2, p_3)$, and $d(p_1, p_3)$, see Fig. 6 (a). The probabilities of distances $d(p_1, p_2)$ vs. $d(p_2, p_3)$ and $d(p_1, p_2)$ vs. $d(p_1, p_3)$ in consecutively occuring position changes are visualized in Fig. 6 (b) and (c), respectively. These probabilities reveal the following peculiarities which cannot be expected to occur by chance. First, note that the diagonal in Fig. 6 (b) and the first column in Fig. 6 (c) are overrepresented. This means that after a position change there is a tendency to have a position change of the same distance, and *this change brings the player back to the previous position.* In other words, there is a tendency to return to the previously visited sector, regardless of the travelled distance. Second, high probabilities are accumulated in the proximity of the diagonal in Fig. 6 (c); for $d(p_1, p_2) > 6$ values are lower between the first column and the diagonal than *on* the first column and the diagonal. Therefore, if a player did not immediately return to the previous sector after a long-distance movement, she is likely to move to the vicinity of the new sector. Further note that right to the diagonal in Fig. 6 (b) values are very low, meaning that a position change is unlikely to be followed by a position change with a bigger distance. This is due to the boundedness of the game universe.

These insights motivate us to modify the biased heterogeneous random walkers by adding one more property: Every timestep, a walker randomly decides with probability q to jump back to the previously visited sector, or with probability $(1 - q)$ to follow the behavior defined for heterogeneous random walkers above, where $0 \leq q \leq 1$. A value $q = 0$ recovers the previous version of the walkers, $q = 1$ leads to walkers constantly swapping between their first and second positions. The last row of Fig. 5 shows the properties for these kind of random walkers when $q = 0.8$. Here the distribution of biweekly jumps is almost identical to the players, entropies are also much more similar than for the case $q = 0$ which is represented in the third row of the figure. However, because of an overemphasis of large jumps (walkers tend to be caught in "loops" going back and forth between the same two sectors), the distribution of daily position distances is now slightly smeared out, see Fig. 5 (m).

Discussion

We analyzed data of daily position changes of 1991 players active over 500 days in the virtual universe of a massive multiplayer online game. This universe has the topology of a network, with properties similar to a lattice or a planar/euclidian network. Analysis of the mobility data and comparison with different types of random walkers simulated on the same network revealed the two essential ingredients which explain the mobility patterns of the players most accurately: Heterogeneity and a tendency to return to recently visited locations. The first observed feature, heterogeneity, revealed itself in the unusually broad distribution of spatial entropies. On the other hand, the distribution of number of unique locations visited, and its logarithm, the random entropy, was found to be much more narrow than expected by chance. This hinted towards a small number of special locations which players prefer to visit (and others which they tend to avoid). The tendency to return to recently visited locations was measured directly, by comparing distances of consecutive position changes.

Comparison with mobile phone data. The cardinal question that remains to be discussed is the following: "How do mobility patterns of human-controlled, virtual characters in an online game correspond with mobility patterns of humans in real life world?" More poignantly, "Is there a real-world benefit in analyzing movement of players in a game?"

By comparing our results with mobility patterns from mobile phone studies, we can affirm the last question. Confer [9], who studied the mobility patterns of 50,000 mobile phone users in a limited area over a certain time interval. They encountered similar biases as we did, e.g. an incompleteness of data in time where a measurement could only be taken when a mobile phone was used. Our measured distributions for random and spatial entropy turn out to be *practically identical* to [9], both in shape and values. Compare Fig. 5 (c) and (d) in this paper with the S^{rand} and S^{unc} distributions of Fig. 2 (A) in [9], respectively. Maxima occur both around 6 and 3.5, respectively (note that the heights of the maxima cannot be compared since y-axes in a probability distribution have arbitrary units depending on bin size). The shapes of the curves are very similar, with a broader and more symmetric distribution for spatial than random entropy. If this finding is not a coincidence or an artefact due to similar sampling, it seems to suggest that certain fundamental behavior patterns of humans can be encountered indepent of the medium they use. While Pardus may be "just a game" to the uninformed observer, constraints and life in the game my be similar to certain aspects in real life world. Therefore it can be possible that specific human behavior patterns in the game are similar.

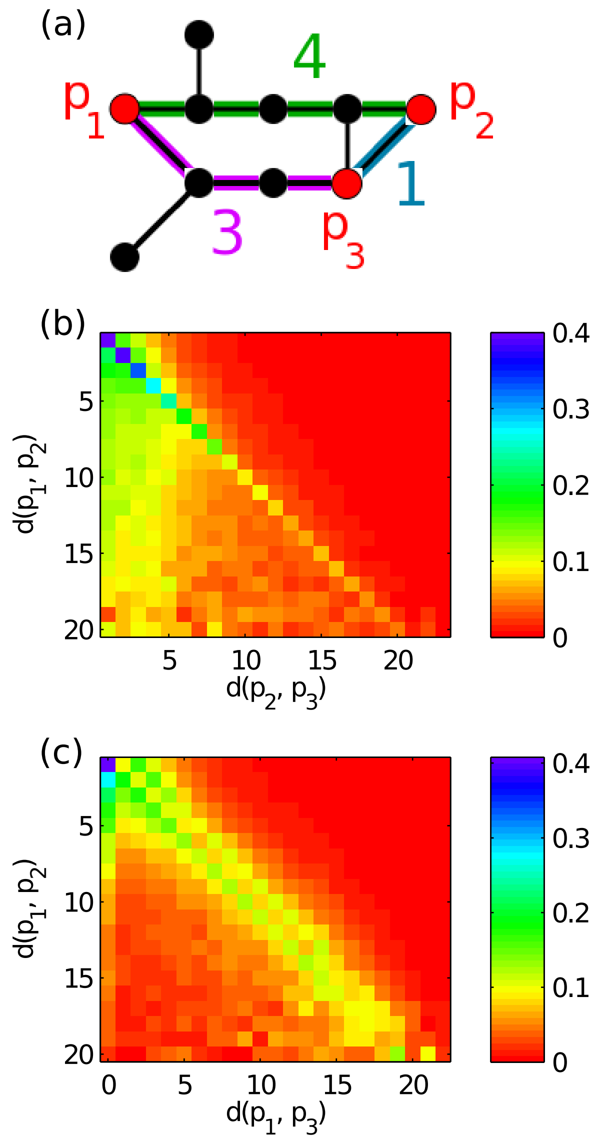

Fig. 6. (a) Two subsequent changes in a player's position, first from a sector p_1 to p_2, then from p_2 to p_3. Shortest path distances between p_1, p_2 and p_3 are measured (in this example, $d(p_1, p_2) = 4$, $d(p_2, p_3) = 1$, and $d(p_1, p_3) = 3$). Probabilities of distances occuring between sectors (b) p_1 and p_2 vs. p_2 and p_3, and (c) p_1 and p_2 vs. p_1 and p_3, for all players and all subsequent changes in positions. There is a tendency to immediately return to the previously visited sector regardless of distance (i.e. often $p_1 \equiv p_3$), visualized by both the diagonal $d(p_1, p_2) = d(p_2, p_3)$ in (b) and the first column $d(p_1, p_3) = 0$ in (c) being overrepresented. Values are normalized, i.e. the sum of all values in each row equals 1.

Above arguments are reinforced by several findings in earlier studies on this online game, e.g. evidence on sociological hypotheses such as the Weak Ties Hypothesis of Granovetter [12] or social balance in friendship and enmity [13].

Outlook. Future work may include extraction of more location points through other logs (e.g. trades, attacks) and a subsequent more fine-grained analysis. Friendship and enmity relations can be incorporated, to track (as)similar mobility behavior occuring between players who stand in any of the possible relationships to each other [13]. Further, the creation and removal dynamics of these relations and their correspondence with co-location or similarity/synchronicity of movement patters could be analyzed.

Materials and Methods

Universe. We have all field-id-to-sector-id mappings and the adjacency matrix of the Pardus universe available; construction of a distance matrix by using a shortest-paths algorithm (Dijkstra) is straightforward. We count sectors which are connected by a wormhole or an X-hole as having distance 1, although players have a higher cost of jumping through X-Holes than for jumping through wormholes. Further, we ignore the fact that the distribution of fields per sector and the locations of wormholes within sectors is not uniform, i.e. hopping from sector to sector has variable costs.

Player positions. We focus on one of the three Pardus universes, *Artemis*. For this universe, we have player mobility data available from day 1 to day 856 of its existence. Players who are inactive for 120 days are automatically deleted. To make sure we only focus on active players, we select all players who have characters in the game between days 236 and 856, and consider the time-span between days 236 and 736 (120 days before day 856). This selects 1991 players active over a time-period of 500 days (data is missing on two days, so we have 498 datapoints in time available). The field-id of these players is logged every day at 05:35 GMT. From these field-ids, we can reconstruct corresponding sector-ids, i.e. the player positions on the universe network's nodes. Note that we do not have full timeseries (down to every second) of positions available, only daily snapshots. This introduces a bias, by which this mobility data should be interpreted as 'long-scale' or 'migratory' movement information, as e.g. a position distance of zero between two consecutive days can either mean standstill, or a movement along an arbitrary number of sectors on that day and by 05:35 GMT a return to the position which was occupied at 05:35 GM the previous day.

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

Action Points. Every game action carried out by a player (trade, movement, attack, etc.) incurs a certain amount of so-called *Action Points* (APs). These points can not exceed a maximum value (this maximum is by default at 5000 but can be pushed by players up to over 10,000). For players having less APs than the maximum, every few minutes a small number of APs is replenished automatically. Once a player's character is out of APs, she has to wait for being able to play on. As a result the typical Pardus player logs in once a day to spend all her APs on several activities within a few minutes. For more details on the specific setup of Pardus see [12, 13].

Random and spatial entropy. As there are 400 sectors in the universe, there is exactly one of 400 distinct positions a player may occupy at a given point in time. We label the total number of unique sectors visited by a player j within all available 498 days as $U(j)$. This number can take values from 1 to 400. Following [9], we call the binary logarithm of $U(j)$ the *random entropy* $H^{\mathrm{rand}}(j)$:

$$H^{\mathrm{rand}}(j) = \log_2 U(j) \qquad [2]$$

We define the spatial entropy $H^{\mathrm{spat}}(j)$ of a player j as

$$H^{\mathrm{spat}}(j) = -\sum_{i=1}^{400} p_i(j) \log_2 p_i(j), \qquad [3]$$

where $p_i(j)$ is the measured probability over 498 days that player j has occupied sector i (we follow the convention $0 \log_2 0 = 0$). In the two extreme cases, occupation of a single sector over all 498 days results in a random and spatial entropy of $H^{\mathrm{rand}} = H^{\mathrm{spat}} = -1 \log_2 1 = 0$, while an equally distributed occupation of all sectors results in the maximal possible spatial entropy $H^{\mathrm{spat}} = \log_2 400 \approx 8.6$. This maximum value is reached for the random entropy if all sectors are visited, regardless of distribution. Note that what we call spatial entropy is labelled *temporal-uncorrelated entropy* in [9].

ACKNOWLEDGEMENTS. We thank the Santa Fe Institute for giving the opportunity to work on this project via the Complex Systems Summer School 2010, as well as Roberta Sinatra for helpful discussions.

1. Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-free networks. *Physical review letters* 86:3200–3203.
2. Colizza V, Barrat A, Barthélemy M, Vespignani A (2006) The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences* 103:2015.
3. Tang J, Musolesi M, Mascolo C, Latora V (2009) *Temporal distance metrics for social network analysis* (ACM), pp 31–36.
4. Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. *Nature* 439:462–465.
5. Eagle N, Pentland A (2009) Eigenbehaviors: Identifying structure in routine. *Behavioral ecology and sociobiology* 63:1057–1066.
6. González M, Hidalgo C, Barabási A (2008) Understanding individual human mobility patterns. *Nature* 453:779–782.
7. Lee K, Hong S, Kim S, Rhee I, Chong S (2009) *Slaw: A new mobility model for human walks* (IEEE), pp 855–863.
8. Phithakkitnukoon S, Husna H, Dantu R (2008) Behavioral entropy of a cellular phone user. *Social Computing, Behavioral Modeling, and Prediction* pp 160–167.
9. Song C, Qu Z, Blumm N, Barabasi A (2010) Limits of predictability in human mobility. *Science* 327:1018.
10. Lambiotte R, et al. (2008) Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and its Applications* 387:5317–5325.
11. Cattuto C, et al. (2010) Dynamics of person-to-person interactions from distributed rfid sensor networks. *PloS one* 5:e11596.
12. Szell M, Thurner S (2010) Measuring social dynamics in a massive multiplayer online game. *Social Networks* in press.
13. Szell M, Lambiotte R, Thurner S (2010) Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences* 107:13636–13641.
14. Castronova E (2006) On the research value of large games. *Games and Culture* 1:163–186.
15. Bainbridge W (2007) The scientific research potential of virtual worlds. *Science* 317:472.
16. Bartle R (2004) *Designing virtual worlds* (New Riders Games).
17. Castronova E (2005) *Synthetic Worlds: The Business and Culture of Online Games* (University of Chicago Press, Chicago).
18. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D (2006) Complex networks: structure and dynamics. *Physics Reports* 424:175–308.
19. Dorogovtsev S, Mendes J (2003) *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press).
20. Latora V, Marchiori M (2001) Efficient behavior of small-world networks. *Physical Review Letters* 87:198701.
21. Noh J, Rieger H (2004) Random walks on complex networks. *Physical Review Letters* 92.