

Fundamental structures of dynamic social networks

Supplementary Information Appendix

Vedran Sekara, Arkadiusz Stopczynski & Sune Lehmann

August 12, 2016

Contents

S1 Summary of main results	5
S2 Data	7
S2.1 Construction of temporal network slices	7
S2.2 Selection of time-scales	8
S3 Gatherings	9
S3.1 Detecting gatherings	9
S3.1.1 Partitioning the dendrogram	9
S3.1.2 Temporal decay function	10
S3.1.3 Gathering timescales	12
S3.2 On- & off-campus gatherings	13
S3.3 Gathering statistics	13
S3.4 Temporal communities	15
S3.4.1 Optimal clustering partition	18
S3.5 Dyadic relations	19
S3.5.1 Dyad statistics	20
S4 Cores	20
S4.1 Extracting cores	20
S4.2 Core statistics	21
S4.3 Subcores	23
S4.4 Work & recreational cores	23
S4.5 Meeting regularity	25
S4.6 Ego viewpoint	25
S5 Predicting behavior from routine	25
S5.1 Comparing with previous studies	26
S5.1.1 Difference in populations	27
S5.1.2 Geospatial resolution	27
S5.1.3 Effects of binning	27
S5.2 Data for prediction	28
S5.2.1 Social vocabulary	28
S5.2.2 Location vocabulary	29
S5.2.3 Convergence of states	29
S5.3 Prediction	29
S5.4 Temporal aspect of predictability	30
S6 Social prediction	32
S6.1 Null models	32
S6.2 Comparison	32

S7 A dynamic random geometric graph model	33
S7.1 Single time-slice RGG parameters	33
S7.2 Model versus the real world	34
S7.3 A dynamic RGG model	35
S7.4 Random walks in 2D	36
S7.5 The dynamic RGG model does not generate recurring meetings	36
S7.6 Summary	37
S8 Coordination of meetings	38

List of Figures

S1	Summary of main findings	5
S2	Fundamental difference induced by temporality	6
S3	Correlation between network slices	8
S4	Temporal coupling	10
S5	Illustration of group dendrogram	11
S6	Stability measures of gatherings	11
S7	Effect of decay parameter on robustness	12
S8	Gathering micro fluctuations	13
S9	Statistical features of gatherings	14
S10	Duration as function of gathering size	15
S11	Gathering stability versus size and duration	16
S12	Temporal patterns of gatherings	16
S13	Gathering participation profile	18
S14	Estimating optimal number of clusters using the Gap measure	19
S15	Statistical features of dyadic gatherings	20
S16	Extracting cores	21
S17	Core selection boundary	22
S18	Core statistics	22
S19	Number of cores per individual	22
S20	Subcore structures and statistics	23
S21	Distinction between work & recreational cores	24
S22	Meeting regularity of cores	25
S23	Ego-centric perspective of cores	26
S24	Cell tower resolution	27
S25	Effects of binning on predictability	28
S26	Time-series of states	29
S27	Distribution of the number of distinct states	30
S28	Distribution of entropy and predictability	30
S29	Correlation between social and location entropy	31
S30	Temporal aspects of predictability	31
S31	Social prediction	33
S32	The random geometric graph model	34
S33	Dynamic random geometric graph model	35
S34	Increased coordination prior to a meeting	39

List of Tables

S1	Data overview	7
----	-------------------------	---

S1 Summary of main results

Figure S1 illustrates our main findings. Social groups display a complex temporal behavior, with dynamics spanning multiple time-scales. Typically, incorporating the temporal dimension drastically complicates the mathematical description of complex networks such that community detection methods require sophisticated mathematical heuristics to disentangle the web of interactions. By observing social interactions at the right time scale—when the temporal granularity is higher than the turnover rate—we can directly observe social gatherings. Figure S1a-c shows social networks obtained using three temporal-windows of increasing size. While daily and hourly windows of aggregation obscure social relations (Fig. S1a-b), a micro-level description directly reveals the fundamental structures. Applying a simple mathematical matching scheme across time-slices reveals dynamically evolving gatherings with soft boundaries and stable cores (Fig. S1d). Unlike the typical community detection assumption of binary assignment, it is clear that some members participate for the total duration of the gathering, while others only participate briefly (Fig. S1d). Matching cores across longer time-scales allows us to observe dynamics that unfold over weeks and months. Cores provide a strong simplification of the social dynamics (Fig. S1e), and are manifested throughout other data channels such as coordination behavior via call and text messages. To demonstrate the saliency of our description we use the social contexts provided by cores to quantify the predictability of social life and give a proof of concept of a new type of non-routine prediction.

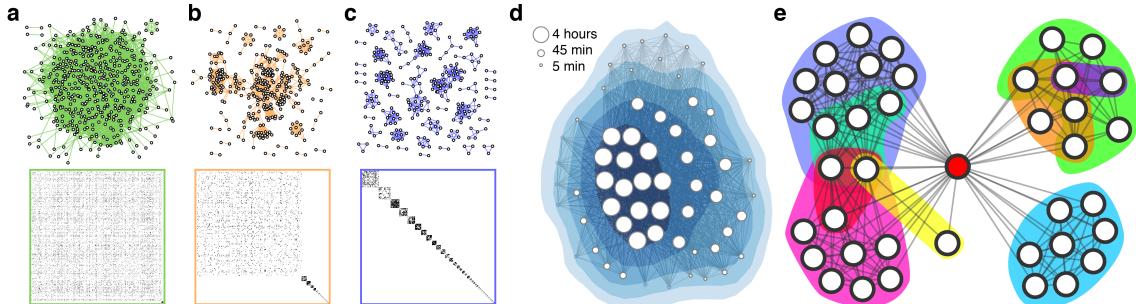


Figure S1: Summary of main findings. **a-c**, Network slices obtained by slicing the social dynamics using varying temporal windows (1 day (green), 1 hour (orange), and 5 minutes (blue)). Below, adjacency matrices colored in agreement with networks, and sorted according to component size. **d**, Gatherings have soft boundaries. The size of each node represents the level of participation. **e**, Cores simplify social dynamics and provide a context for social interactions.

A note on correlations induced by temporality

It is instructive to consider the magnitude of the correlations induced by temporality. We investigate this question as follows. Let us say that the network in Fig. S1c has E_c edges. We now generate a random network based on Fig. S1a, where E_c edges are sampled at random. Now we compare the original time-slice (Fig. S1c) with this reference network. Fig. S2 summarizes the key differences. Figure S2A and B provide a visual comparison of the networks, which strikingly illustrates how far from a random sample of the aggregated

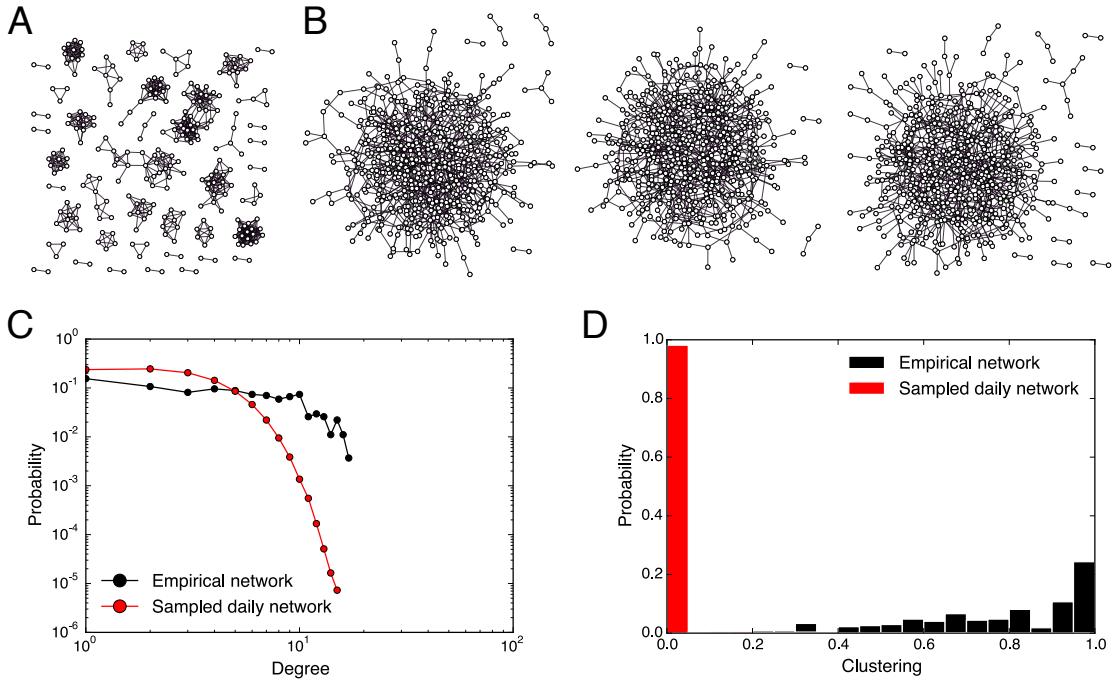


Figure S2: Fundamental difference induced by temporality. **A.** A snapshot of social interactions within a 5-minute temporal bin, with E edges. **B.** Three reference networks constructed by randomly sampling E edges from the daily interaction graph (Fig 1A, main manuscript). **C.** Degree distribution for the network in **A** and 1000 random references networks constructed as described in **B**. **D.** Clustering distribution. Sampling E edges from the daily network produces mainly one big component with low probability of observing high degree nodes and with low clustering as compared to the temporal snapshot.

network each temporal network slice is. Subfigure **A** shows the original network from Figure S1c with E_c edges, and Fig. S2 shows three examples of subsamples of E_c edges sampled from the network in Figure S1a. We see that whereas the original time-slice has many medium sized components with high clustering, the resampled networks form a single, sparse component with low clustering and a few isolated dyads. The basic network statistics confirm how radical this change is. Figure S2C shows the change in degree distribution, and Figure S2D quantifies the change in the clustering-distribution as we go from the highly clustered real-world data to the tree-like random subsamples.

— ● —

The remainder of the SI document is organized as follows. Section 2 describes the dataset, Section 3 explains the details of how gatherings are constructed, and describes their basic statistics, including a discussion of dyads. Section 4 shows how cores are extracted and goes into detail on their temporal behavior, as well as the sub-core structures. In Section 5 we provide full details regarding the routine-based geospatial prediction as well as the social context prediction presented in the main text. Section 6 provides background on our

model for purely social prediction and, finally, Section 7 discusses background information regarding the coordination leading up to meetings.

S2 Data

We consider a dataset from the *Copenhagen Networks Study*. It spans multiple years and measures with high resolution: physical interactions, telecommunications, online social networks, and geographical location. In addition, the dataset contains background information on all participants (personality, demographics, health, politics). These data are collected for a densely connected population of approximately 1 000 students at a large European university. Data is collected by running custom built applications installed on 1 000 smartphones (Google Nexus 4). Full details can be found in Ref. [1].

In this manuscript we focus on detecting and tracking co-located groups of individuals during a representative period of five months (roughly one semester), collected between January 1st and June 1st of 2014. The Bluetooth sensor collects proximity data ($\sim 0 - 10$ m) of the form (i, j, t, s) , where each interaction implies that person j has been in proximity of person i at time t , where the devices observe each other with signal strength s [2]. Bluetooth scans do not constitute a perfect proxy for face-to-face interactions. In fact multiple scenarios exist where people in close proximity do not interact and vice versa, nevertheless Bluetooth can successfully be applied in order to sense social networks [1–3]. Further, our gathering/core-description naturally filters our spurious connections by considering social structures that occur over across extended periods of time. Gatherings and cores are identified in the proximity network, and the remaining communication channels are used for validation purposes. In addition, we reserve proximity data from the month of May for validation purposes. Table S1 shows statistics across the various data sources for 814 individuals, on whom we focus due to their high data quality.

Data source	Total	Unique
Bluetooth interactions	14 673 869	154 818
Call & text interactions	75 364	1 216
Geographic locations	18 603 072	-
WiFi access points	1 663 483 977	2 412 702

Table S1: Data overview from January 1st – June 1st. Bluetooth and call & text logs are summarized for within-participant relations and do not include external interactions. The unique field denotes the number of distinct observation of each quantity, e.g. number of uniquely observed links.

S2.1 Construction of temporal network slices

The data collection application triggers Bluetooth scans every five minutes from the time a phone is turned on; for this reason, the collection of sensors does not follow a global schedule. To account for this behavior we divide all temporal information into absolute time-windows, Δ minutes wide. Within each temporal bin, we draw a unweighted undirected link between two individuals if either i has seen j or vice versa.

S2.2 Selection of time-scales

The scanning frequency of the application sets a natural lower limit of the network resolution to 5 minutes, however, there is no such upper limit for aggregation. So-called natural timescales have previously been investigated for specific networks with respect to their global topological properties [4–6]. Here we consider the correlation between slices (or turnover of nodes between slices) defined as $C = |E_i^\Delta \cap E_{i+1}^\Delta| / |E_i^\Delta \cup E_{i+1}^\Delta|$, where E_i^Δ denotes the set of edges that are observed in bin i with given temporal width Δ . According to Fig. S3 the average correlation decreases sharply as function of window-size, achieving maximum correlation for small bins. Slicing network dynamics into short slices (high resolution) also disentangles the network (Fig. S1), thus when time slices are shorter than the group’s turnover rate, we can directly observe individuals’ group affiliations. Based on Fig. S3 we chose a temporal width of 5 minutes, but windows of 10 minutes could have been chosen without deterioration of results.

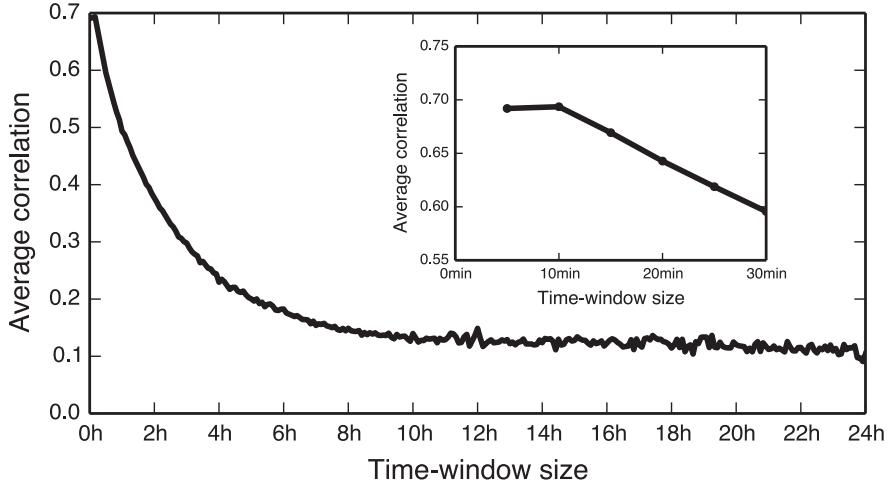


Figure S3: Average correlation between network slices, averaged across all time-bins. Inset shows a closeup for the smallest bin-sizes. Calculated for proximity data form March 2014.

When we consider the duration of time-windows, there are two aspects to consider, a *methodological perspective* and a *practical perspective*. From the methodological perspective, using any finite time window shorter than 10 minutes (as argued in Fig. S3), will produce analyses and results that are consistent with what we have presented in the main text. Using a resolution finer than 5 minutes could potentially be interesting if one was interested in measuring the precise dynamics of group formation (answering questions such as: what is the precise sequence of arrivals?, does a certain group member consistently arrive before everyone else? etc.) or group dissolution (e.g. someone leaving 30 seconds before the rest of the group could be an important social signal).

Considering the practical perspective, it is clear that measuring e.g. once every millisecond would produce 300 000 times more data than we are currently collecting, while adding very little extra information about the social system. For the type of dynamics

considered in the main text (e.g. individuals moving from one gathering to another) 5 or 10 minute time-bins arguably present an useful trade-off between measurement accuracy and the meaningful network changes.

S3 Gatherings

In this section we describe how connected components in the proximity network are matched across short timescales into dynamical ensembles, which we denote as gatherings. We then present fundamental statistics on gatherings (size distribution, durations, stability, start/end times), and analyze these properties in the light of on/off-campus behavior. Finally, we focus on identifying repeated gatherings across longer timescales to infer dynamical communities.

S3.1 Detecting gatherings

In each temporal slice we identify connected components, i.e. nodes that are in close physical proximity as social groups. Since dyadic relationships qualitatively differ from group relations [7, 8] we generally treat components of size two separately.

A gathering is defined as a group that is persistent across time. To identify gatherings we apply agglomerative hierarchical matching, a widely used method, that merges groups based on their mutual distance (defined below) [9, 10], using a matching strategy similar to Green *et al.* [11]. Each group is initially assigned to its own cluster, then every iteration-step merges the two clusters with smallest distance according to the single linkage criteria ($\min(d_i(c_t, c_{t'}))$). This merge criterion is chosen because it is strictly local and will agglomerate clusters into chains, a preferable effect when clustering groups across time. The clustering process is repeated until all groups have been merged into a single cluster. Distance between groups is calculated using a modified version of the Jaccard similarity:

$$d(c_t, c_{t'}) = 1 - \frac{|c_t \cap c_{t'}|}{|c_t \cup c_{t'}|} f(\Delta t, \gamma), \quad (\text{S1})$$

where $f(\Delta t, \gamma)$ is a term that denotes the coupling between temporal slices and $\Delta t = t' - t$ denotes the temporal distance between two bins (for consecutive bins $\Delta t = 1$). The function can assume any form, increasing or decreasing; we utilize it to model decay between temporal slices, with the two most prominent forms being: exponential ($\exp(-\gamma(\Delta t - 1))$), and power-law ($\Delta t^{-\gamma}$), see Fig. S4. Thus, by definition the term assumes the value 1 (zero decay) between two consecutive temporal slices. For computational reasons we only focus on gatherings identified using exponential decay with $\gamma = 0.4$, other decay parameters yield similar results see SI Section S3.1.2.

S3.1.1 Partitioning the dendrogram

The method described above iteratively constructs a dendrogram where temporally localized groups are hierarchically clustered. To extract meaningful social structures we need to partition the dendrogram (Fig. S5). Modularity and partition density have previously been applied for similar purposes [12, 13], but these do not generalize well for temporal processes. Instead we consider the cluster stability with respect to local and global measures. Local

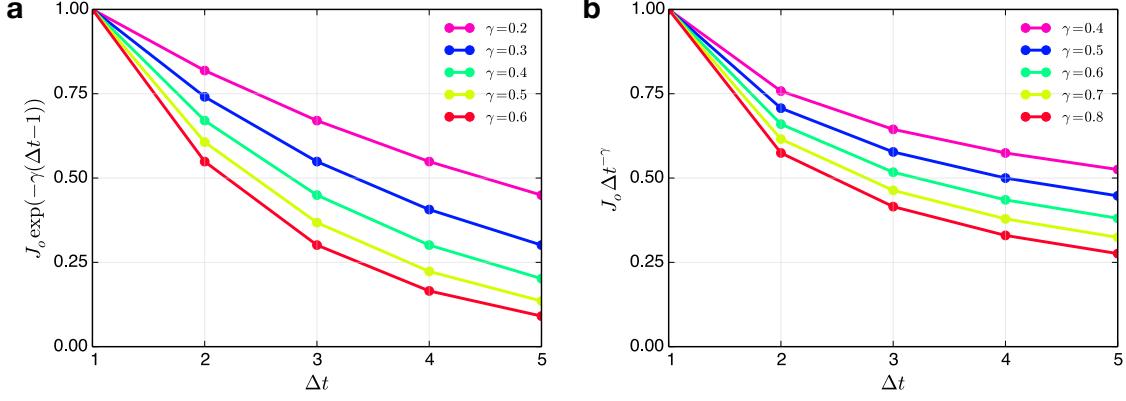


Figure S4: Temporal coupling between two temporal slices as function of the decay parameter γ . By definition the decay is zero between consecutive slices ($\Delta t = 1$). **a**, Exponential decay. **b**, Power law decay.

stability (η) is calculated from the average node-wise overlap between consecutive slices (Fig. S6a), while global (σ) is calculated from the average overlap between all slices and the aggregated structure (Fig. S6b):

$$\eta = \frac{\sum_{t=t_{\text{birth}}}^{t_{\text{death}}-1} J(g_t, g_{t+1})}{t_{\text{death}} - t_{\text{birth}} - 1}, \quad (\text{S2})$$

$$\sigma = \frac{\sum_{t=t_{\text{birth}}}^{t_{\text{death}}} J(g_t, G)}{t_{\text{death}} - t_{\text{birth}}}, \quad (\text{S3})$$

where t_{birth} and t_{death} are respectively the birth and death of the gathering, g_t is a temporal slice, $G = g_{\text{birth}} \cup g_{\text{birth}+1} \cup \dots \cup g_{\text{death}}$ is the aggregated structure, and J is the overlap ($J = |i \cap j| / |i \cup j|$), defined as zero if the gathering has only existed for one time bin. Palla *et al.* [14] applied a related measure to estimate the stationarity of communities. Varying the partition threshold (Fig. S6c) we observe a maximum in both measures, indicating a regime where gatherings are both temporally and globally stable. Threshold values of $d \geq 1/2$ are, however, problematic since they merge gatherings that split into two equally sized parts together with both parts, or vice versa. In this scenario, we find that the desirable behavior is to declare the old gathering as ‘dead’ and identify two new gatherings as born. Therefore, to achieve optimal stability and to avoid issues with unwanted merging we partition the dendrogram at $d = 0.49$.

S3.1.2 Temporal decay function

Here we investigate how robust the inferred gatherings are to perturbation of the γ -parameter and using an alternate decay form. To compare two set of gatherings (identified using different γ -values) we calculate the average maximal overlap between individual gatherings

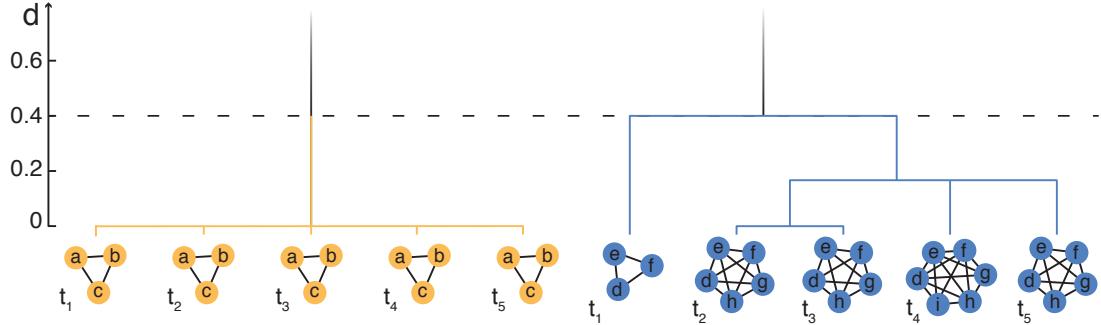


Figure S5: Illustration of constructed dendrogram, depicting distance (d) between groups identified across 5 timebins. The tree is constructed using an exponential decay function with $\gamma = 0.4$. Two gatherings, orange and blue, are inferred by thresholding the tree, where all groups below or equal to the threshold ($d = 0.4$) are merged.

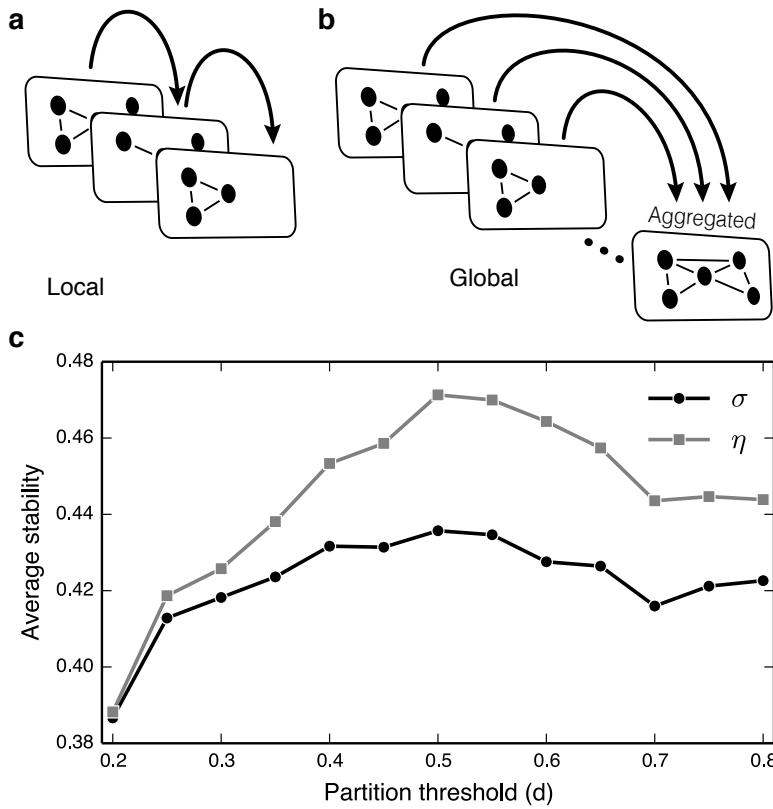


Figure S6: Stability measures of gatherings. **a**, Illustration of the local stability measure, calculated between consecutive slices. **b**, Global measure calculated between each slice and the aggregated structure. **c**, Global (σ) and local (η) stationarity of gatherings as a function of partition value d , averaged over all gatherings identified in January 2014. Gatherings achieve optimal stability around $d \sim 1/2$.

as

$$O_{\gamma\gamma'} = \frac{1}{|G_\gamma|} \sum_{i \in G_\gamma} \max_{j \in G_{\gamma'}} (J(i, j)), \quad (\text{S4})$$

where G_γ denotes the set of gatherings found using a specific value of γ , and $|G_\gamma|$ is the number of gatherings. Overlap is calculated using Jaccard similarity, $J = |i \cap j|/|i \cup j|$, where i and j are respectively gatherings from G_γ and $G_{\gamma'}$. Figure S7 shows the overlap matrix between identified sets of gatherings; in general it assumes overlap values above 0.76 for any choice of parameters. Indicating that the gatherings are robust to even large perturbations.

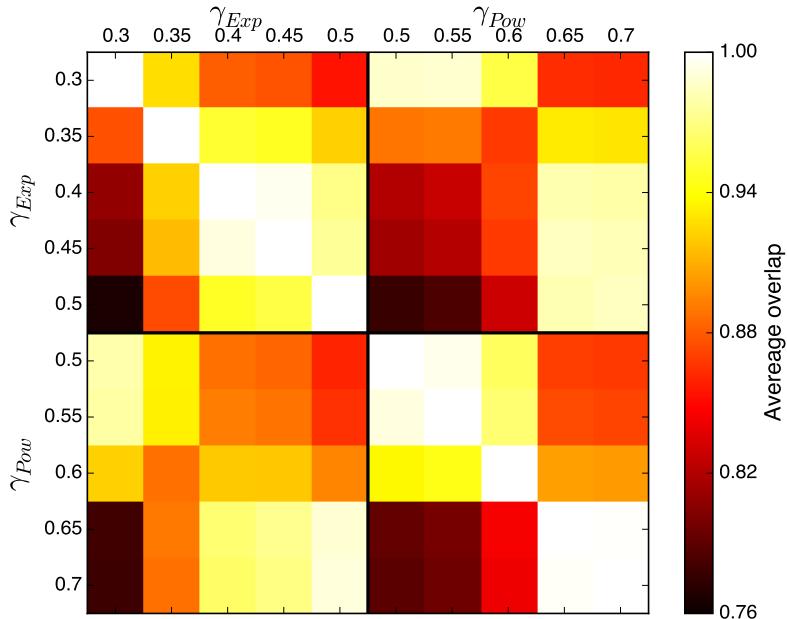


Figure S7: Effect of decay parameter on gathering robustness, calculated using Equation S4. Colorbar shows average overlap between two sets of gatherings, and never drops below 0.76, indicating the robustness of the procedure with respect to parameter perturbations.

S3.1.3 Gathering timescales

The outlined method identifies multiple gatherings, some only exist momentarily while others are sustained for long time periods. One can easily imagine brief encounters between good friends as being more meaningful than prolonged interactions between individuals commuting to work; this therefore raises the question of which meetings are meaningful and which are not.

Here we simply adopt the convention developed by the *Rochester Interaction Record* [15], where meaningful encounters are defined as those lasting 10 minutes or longer. In order to filter out spurious connections, we impose the requirement that a gathering must be observed for at least 4 consecutive time slices to be represented in our statistics.

While dynamics on 20 minutes+ timescales describe the overall evolution of a gathering, micro dynamics on 5-minute scales represent everyday events such as going to the bathroom. Gatherings, therefore, might disappear and reappear within very short time-intervals (Fig. S8a), in such cases we use imputation, see Fig. S8b.

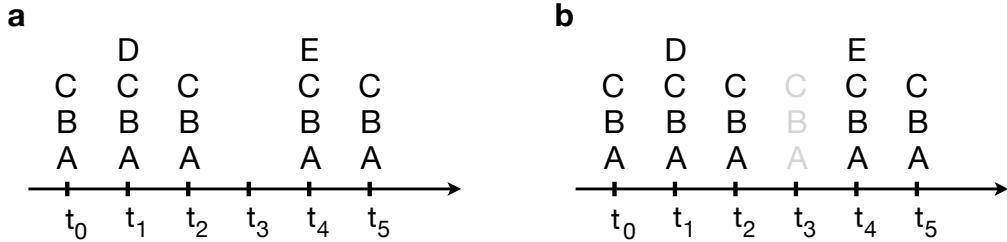


Figure S8: Gathering micro fluctuations. **a**, A gathering appears, disappears, and reappears moments later. **b**, If this occurs in rapid succession we treat the gathering as if it was present in bin t_3 with the same nodes as in bin t_2 .

S3.2 On- & off-campus gatherings

Gatherings are not geographically constrained and therefore free to occur anywhere, in this section we focus on distinguishing between gatherings that occur on- and off-campus. We do this by applying data acquired through the WiFi channel, where each mobile phone scans for nearby wireless network access points (AP) every 5 minutes or less and logs their unique identifier and name. The entire university campus is densely covered by wireless networks and it is therefore highly unlikely that students located on campus will not see an university AP. Because the names of campus APs follow a uncommon and standardized naming scheme we can infer when a student is close to a campus access point and use this as a proxy of being on or off campus. For each participant we construct a 5-minute binned vector containing binary values 0 (not on campus) and 1 (on campus). Because gatherings are an ensembles of people we perform majority voting across nodes for each time-bin to determine whether the gathering was on campus or not in that specific time-bin. Since gatherings are spatio-temporal entities, we also perform a majority voting across all time-bins to achieve a hard split and determine whether a gathering mainly occurred on- or off-campus. This yields 13 872 off- and 9 195 on-campus gatherings and with an 92.84% average agreement between votes. The primary reason for disagreement are mobile gatherings traveling to or from campus (e.g. on a bus or walking).

S3.3 Gathering statistics

Gatherings show a broad distribution in both size and duration, see Fig. S9. Dividing gatherings into on- and off-campus categories reveals that meetings occurring on university campus are larger (Fig. S9c), but have considerably lower probability of lasting longer than 4 hours (Fig. S9d). This suggests that large meetings are mainly driven by the class schedule, while meeting duration is determined by social context. Because meetings occurring outside of campus often require increased coordination, our hypothesis is that once groups meet, they will interact across longer periods of time for the meeting to pay off with respect to the organizational cost.

It is also interesting to consider the duration of each meeting as function of the total number of nodes that participate in it. Figure S10a shows broad distributions of duration across all sizes, however, both the mean and median are quite stable and reveal that small

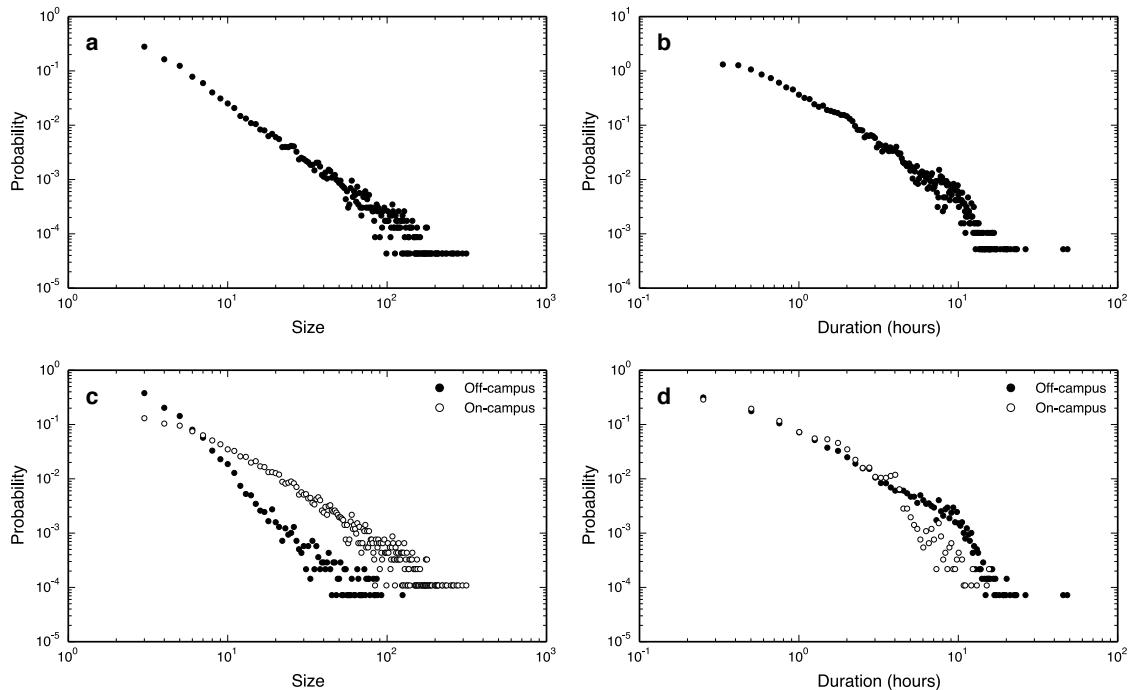


Figure S9: Statistical features of gatherings. Summarizing size and duration distributions for all gatherings (panels a-b) and conditioned on off/on-campus meetings (c-d). **a,** Gathering size distribution. **b,** Gathering duration distribution. **c,** Gathering size distribution for on- and off-campus meetings. **d,** Gathering duration distribution, based on location (on/off-campus).

gatherings on average have shorter durations compared to larger meetings. Further dividing the data into on- and off-campus categories (Fig. S10b) shows that small meetings on and off campus are quite similar with respect to duration, while larger meetings tend to last longer, provided that they occur off-campus.

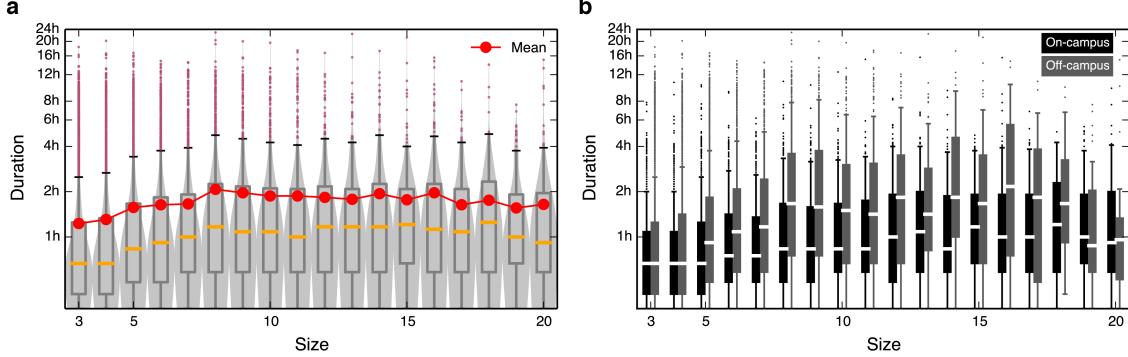


Figure S10: Duration as function of gathering size. **a**, Violin plot shows the distribution of durations as function of size, summarized across all gatherings. **b**, Box plot of the duration distributions divided into on- and off-campus meetings.

Combining the statistics above with Eqs. (S2-S3) we can explore how gathering stability depends on size and duration of meetings. Figure S11a reveals that the local churn between consecutive time slices is quite constant, irrespective of size, indicating that there is a low turnover of nodes between slices—on average much lower than predefined by the partition threshold. For small gatherings, however, we observe finite size effects due to the partition threshold (see sec. S3.1.1). Global stability is lower, but also fairly independent of gathering size. With respect to duration (Fig. S11b), local stability increases as meetings duration increases, revealing that longer meeting have lower turnover of nodes between consecutive slices. The global measure shows similar behavior. It achieves a slightly lower stability and shows that gatherings are globally stable independent of duration. This combination of a constant turnover between slices and a global stability suggests that gatherings contain groups of highly interacting individuals, that are present throughout the entirety of the meeting, while other individuals participate infrequently and are constantly being replaced. Similar social structures have previously been observed in the social science literature where the individuals have been defined as core members [16].

Finally we can look into specific temporal patterns with regards to when on and off campus gatherings occur, i.e. in which 5 minute time-bins they first appear and later disappear. Figure S12a reveals that on-campus gatherings have increased probability of occurring exactly on the hour, while off-campus meetings are evenly distributed—clearly showing the effect of the class schedule. A similar case is seen for the probability of dissolving (Fig. S12b), where on-campus meetings mainly end on integer hour values.

S3.4 Temporal communities

So far each gathering only contains information about its local appearance, to gain a dynamical picture we match gatherings across time. Due to soft boundaries, a strict

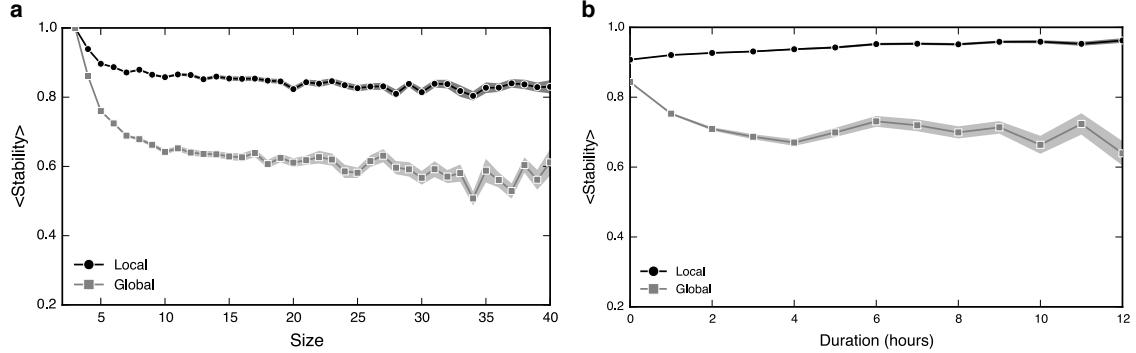


Figure S11: Global and local gathering stability. **a**, Average stability as function of size, averaged across all gatherings with a specific size. Full lines denotes mean, while shaded areas shows the standard deviation of the mean. **b**, Average stability as function of duration, binned into one hour wide bins. Fully drawn lines denote the mean, while shaded areas illustrates its deviation.

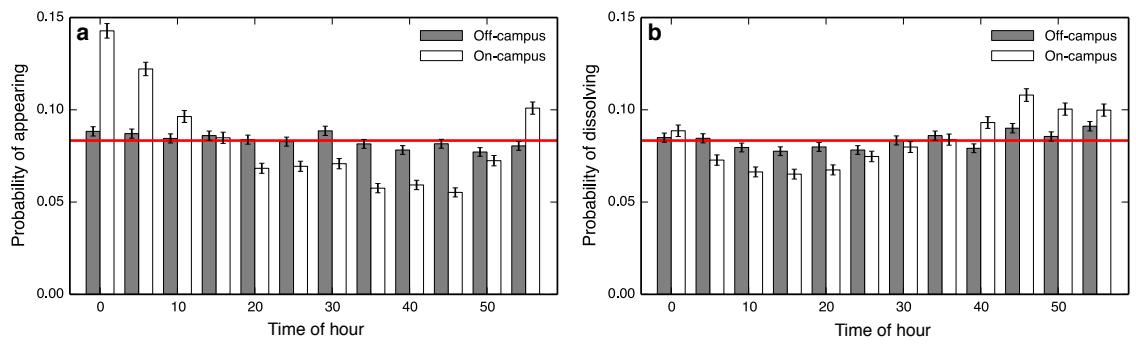


Figure S12: Summary statistics of gathering temporal patterns, summarized for on- and off-campus meetings. Red line denotes the uniform probability distribution for the case where all states are equally probable. **a**, Probability of a gathering appearing, calculated from the first time-bin in which we observe it. **b**, Probability of dissolving, i.e. the last time-bin where we observe a gathering.

matching criteria is not a feasible method, since a person who coincidentally walks past a group might be included in it. Thus, we expect noise to be present in each gathering. To mitigate this effect we instead match gatherings according to the participation levels of their constituent nodes. Counting the fraction of times a nodes has been present in the gathering we construct a normalized participation profile, see Fig. S13a. Again, gatherings are matched according to their individual participation profiles using agglomerative hierarchical clustering. Since nodes, however, no longer assume binary values but may assume participation levels in the interval $0 < n_i \leq 1$, we calculate distance based on a continuous version of the Jaccard similarity:

$$D(G_i, G_j) = 1 - \frac{\sum_{n=1}^N \min(G_i, G_j)}{\sum_{n=1}^N \max(G_i, G_j)}, \quad (\text{S5})$$

where G_i is a vector containing node-wise participation values for gathering i , and N is the total number of nodes in $G_i \cup G_j$. The two functions max and min act piecewise on the two vectors, and $D(G_i, G_i)$ is defined as 1 between two gatherings that have zero overlap. When merging clusters of gatherings (\mathcal{G}) we apply the average linkage criterion and define the average distance between them as

$$D(\mathcal{G}, \mathcal{G}') = \frac{1}{|\mathcal{G}| |\mathcal{G}'|} \sum_{G \in \mathcal{G}} \sum_{G' \in \mathcal{G}'} D(G, G'), \quad (\text{S6})$$

where $|\mathcal{G}|$ denotes the cardinality of a set of gatherings. Other linkage criteria can also be used, such as complete or Ward-linkage. Iteratively this method builds a dendrogram with gatherings as leafs. Thresholding the tree partitions similar gatherings together into communities. A community then consists of all nodes from its constituent gatherings, but it also inherits their individual participation profiles. Thus we need a method to construct a community participation profile from its subcomponents. This can be done using two methods: weighted or unweighted, with the differences illustrated in Fig. S13b. The unweighted method takes into account the gathering participation profiles and calculates the average, weighing each gathering equally:

$$C = \frac{1}{|\mathcal{G}|} \sum_{G \in \mathcal{G}} G. \quad (\text{S7})$$

The weighted version instead assigns each gathering a weight according to its lifetime (τ_{life}), i.e. number of temporal bins it has been present:

$$C^{\text{weighted}} = \frac{1}{\sum \tau_{\text{life}, G}} \sum_{G \in \mathcal{G}} \tau_{\text{life}, G} G. \quad (\text{S8})$$

Both measures comparatively construct similar dynamical communities and yield similar overall statistics; we choose the the weighted version, because it is slightly less influenced by noise.

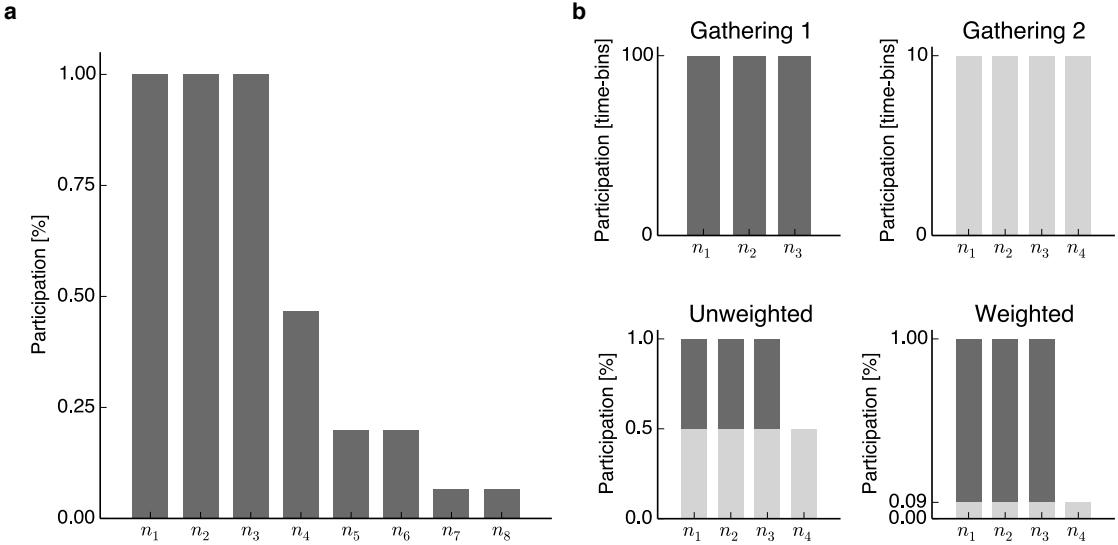


Figure S13: Illustration of gathering participation profiles. **a,** Each vertical bar indicates the overall fraction of time a node has spent in the gathering. Values are normalized according to the total gathering lifetime. **b,** Illustration of how to construct a community participation profile based on its constituent gatherings. Top figures depict two gatherings with unequal lifetimes. Bottom plots illustrate the principles behind the unweighted and weighted methods. Bar colors indicate the role each node plays in the community profile.

S3.4.1 Optimal clustering partition

Applying hierarchical clustering to merge similar gatherings into communities again leaves one open question: Which partition value is the optimal? Using a similar line of reasoning to what we presented in Sec. S3.1.1, we can argue that a threshold value of $D = 0.49$, is preferable. It is, however, possible to estimate the exact optimal threshold value, which we can compare to our hypothesized guess. Applying a heuristic inspired by the Gap statistic introduced by Tibshirani *et al.* [17], we compare the clustering according to a null model distribution (for a comprehensive survey of methods see Milligan and Cooper [18]). Given a total of m gatherings clustered into k clusters (communities): C_1, C_2, \dots, C_k we calculate the within-dispersion measure as,

$$W_k = \sum_{r=1}^k \frac{1}{2|C_r|} \sum_{i,j \in C_r} D_{ij}, \quad (\text{S9})$$

where D_{ij} is defined in equation S5 and denotes the pairwise distance between gatherings i and j that both belong to cluster C_r . Again $|\cdot|$ denotes the cardinality of a cluster, i.e. the number of gatherings that are clustered in C_r . The factor 2 takes double counting into account. Thus W_k is the accumulated within cluster sum of differences around the cluster mean. Applying the principles from Tibshirani *et al.* [17] we compare $\log(W_k)$ to an expected value generated by a null model distribution of the data. The gap measure is

defined as

$$\text{Gap}(k) = 1/B \sum_{b=1}^B \log(W_{kb}) - \log(W_k), \quad (\text{S10})$$

where B denotes the number reference data sets. The optimal number of clusters is then the value of k for which $\log(W_k)$ falls furthest below the reference data curve, i.e. the value of $k = k^*$ such that

$$k^* = \operatorname{argmin}_k \{k | \text{Gap}(k) \geq \text{Gap}(k+1) - \tilde{s}_{k+1}\}, \quad (\text{S11})$$

where $\tilde{s}_k = s_k \sqrt{(1+1/B)}$ and s_k is the standard deviation of $\log(W_{kb})$ over the B synthetic datasets. Each null model is constructed by assigning random participation values, chosen from a uniform distribution, to random nodes, thus creating reference gatherings with similar size distributions. According to Fig. S14 the gap statistic achieves a minimum, indicating that the optimal place to cut the dendrogram is at distances of $D = 0.50$, in good agreement with the previously stipulated value. A threshold value of $D = 0.50$ is, however, problematic since it will cluster gatherings with 50% overlap. To avoid this issue we cut our dendrogram at $D = 0.49$, merging 23 067 gatherings into 7 320 distinct dynamic communities.

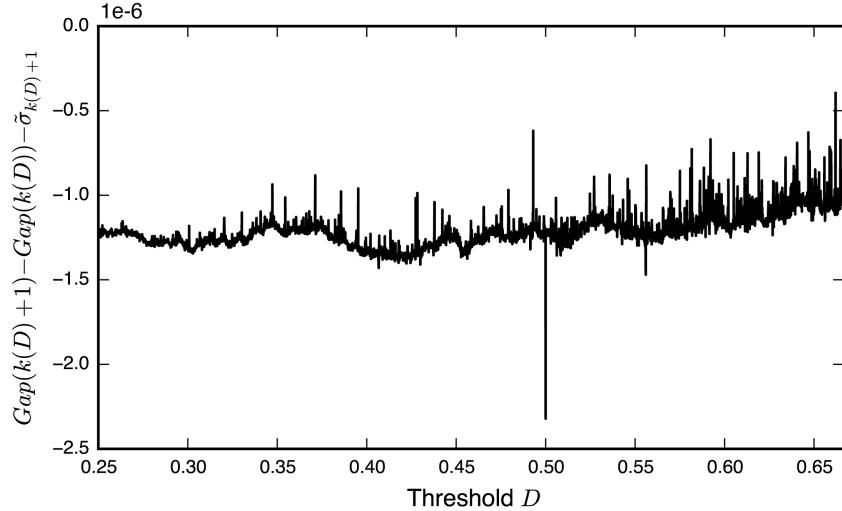


Figure S14: Estimating the optimal number of clusters using the gap measure, with the number of clusters k being directly related to the threshold value D .

S3.5 Dyadic relations

Dyads are a special case of social relations and are hypothesized to be qualitatively different from groups [7, 8], but the way we identify them is similar to the outlined framework in Section S3.1. Since dyads can only be components of size two, tracking their evolution is trivial, because we can apply a strict merge criterion, i.e. require 100% overlap. This

also applies in the case when tracking repeated appearances (dyadic cores) across the full duration of the dataset. Thus, the only distinction between a gathering and a dyadic gathering lies in the fact that size of the gathering will always be fixed, meaning if a dyad evolves into a group (of any size larger than two) then we claim that the dyad in question has ended and a new group relation has been initiated. In accordance with gatherings, we require dyadic gatherings to be present for at least four consecutive bins (20 minutes) to be considered as meaningful. In total we observe 34996 dyadic gatherings and 4844 unique dyads.

S3.5.1 Dyad statistics

Similar to group relations, dyadic gatherings produce a broad distribution of durations comparable to the gatherings of groups (Fig. S15a). In addition, dyads also produce a broad distribution of repeated appearances, see Fig. S15b, with a majority of gatherings only being observed once, while others on average can appear multiple times per day.

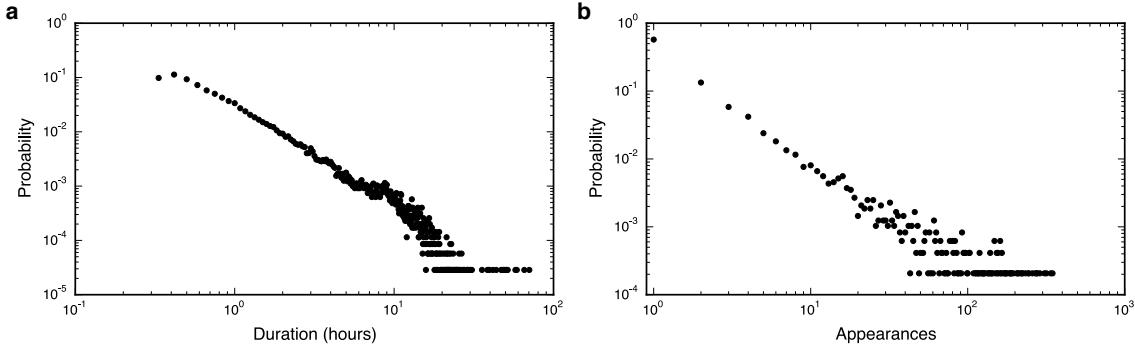


Figure S15: Statistical features of dyadic gatherings. **a**, The distribution of duration for dyadic gatherings. **b**, The distribution of the number of appearances.

S4 Cores

In the previous section (see Fig. S11b) we found that certain individuals are present for a majority of a gathering's lifetime. In this section, we begin by describing how such *cores* are extracted from temporal communities. Then, we describe the fundamental statistics of cores (distribution of appearances, distribution of sizes, individual membership in cores, subcore-statistics), as well as the key differences between work- and recreational cores. Finally, we study the temporal patterns and entropy of core-appearances.

S4.1 Extracting cores

Nodes within each community have varying attendance (see Fig. S13a), some are only members for a limited time, while others interact over extended periods of time. Thus, participation profiles show pronounced core structures, highlighting individuals that act as ‘generators’ of each community.

Consider the participation levels of individuals as ordered profiles (Fig. S16a), where each bar denotes the fraction of time a node has spent in a community relative to the community's total lifetime. A significant gap in this profile identifies core nodes. We compare this to a participation profile generated by a random process (Fig. S16b), where we pick a random participation level between 0 and 1 (for each node) from a uniform distribution. The maximal gap in this random profile thus tells us whether the real gap is significant. We can estimate the average expected gap size and deviation by generating many ($N = 10\,000$) random participation profiles. Generalizing this notion to all sizes of communities we evaluate how significant a gap is compared to the expected value generated at random. The decision boundary in Fig. S17 divides gap sizes into two regions. If the actual gap is greater than the average null-model gap μ_{random} plus one standard deviation σ_{random} , we define the core to be significant. Thus, we only keep cores with gap sizes above $\mu_{\text{random}} + \sigma_{\text{random}}$. According to this criterion, we find that 7 146 out of the 7 320 (97.6%) inferred communities display a pronounced core structure.

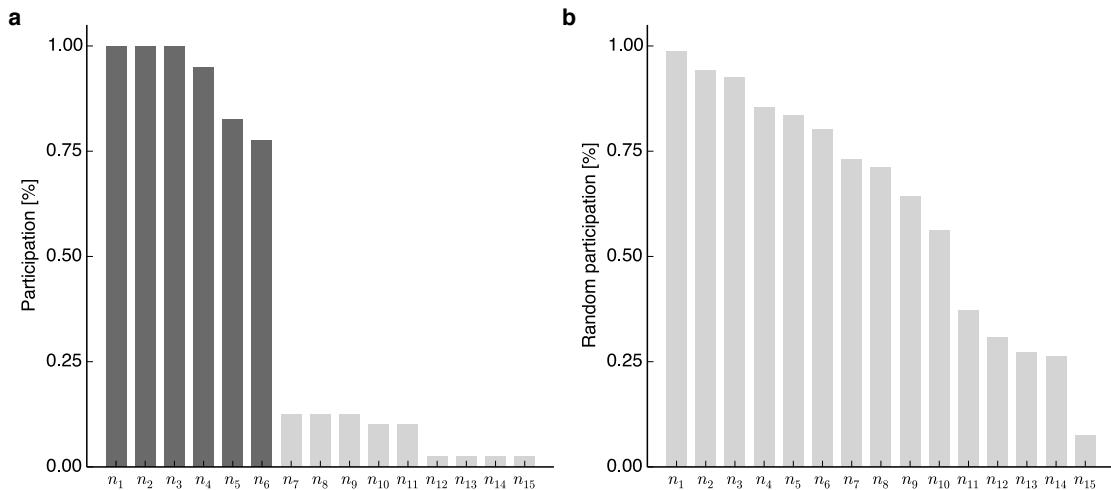


Figure S16: Extracting cores from community participation profiles. Dark gray bars denote nodes with participation levels above the maximal gap. **a**, Ordered participation profile for a community composed of 15 individuals. **b**, Similar as in panel a but generated from a uniform random distribution.

S4.2 Core statistics

Cores have a broad distribution of the number of appearances (Fig. S18a), ranging from cores appearing only once to on average occurring multiple times per day. The size distribution is also heavy-tailed (Fig. S18b).

To produce meaningful temporal statistics we henceforth focus on cores which on average are observed more than once per month, this limits our focus to individuals that appear in these. Fig. S19 shows the distribution of the number of cores each individual is part of; it is a broad distribution with a majority of individuals partaking in few cores while a small minority of users are extraordinary social and have more than 10 cores.

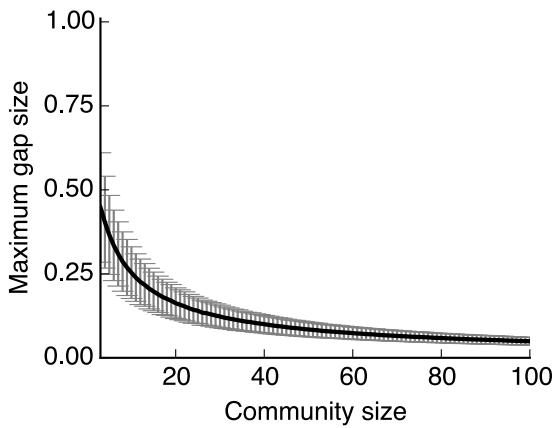


Figure S17: Core selection boundary. Decision boundary calculated from $N = 10\,000$ independent trials for each size. Black line denotes the mean max gap value while error bars indicate the standard deviation.

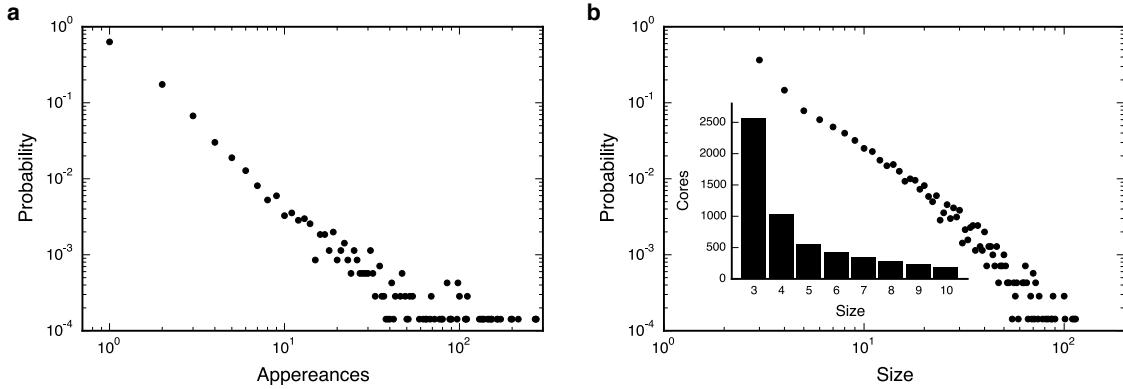


Figure S18: Core statistics, where we focus on cores of size greater than two. **a**, Probability distribution of the number of appearances per core. **b**, Size distribution. Inset shows raw numbers for specific sizes.

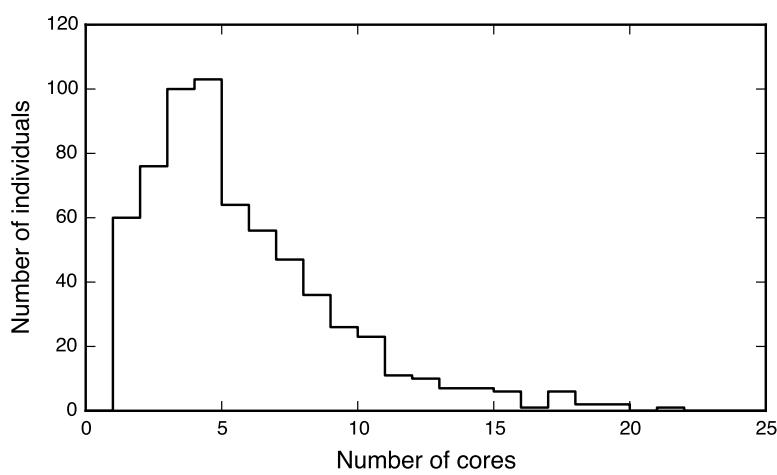


Figure S19: Distribution of the number of cores per individual. Calculated for users that appear in frequently observed cores, i.e. cores that on average are observed more than once per month. A minority of individuals partake in more than 10 cores, with the distribution being centered around a mean value of ~ 5 .

S4.3 Subcores

Because cores span a wide range of sizes small cores can appear as subcores embedded within larger ones, see example in Fig. S20a. In fact, our methodology allows for and identifies highly overlapping and hierarchically stacked structures. We define a core to be a subcore if, and only if, it is fully contained in the larger one. Figure S20b shows a broad distribution in the number of subcores that are contained within individual cores, with a majority of cores only containing one subcore, while other can contain more than ten. There is of course a dependence on size, such that bigger cores have larger probability of containing more subcores. Figure S20c quantifies this phenomenon by considering the fraction of subcores that cores of size s contain.

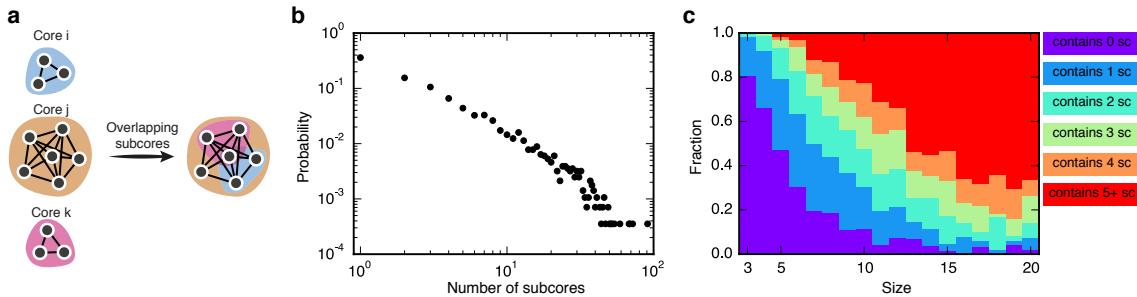


Figure S20: Subcore structures and statistics. **a**, Illustration of overlapping and hierarchically stacked core structures. **b**, The distribution of the number of subcores contained within individual cores. **c**, Fraction of cores that contain 0, 1, 2, 3, 4, and 5+ subcores (sc) as function of core size.

S4.4 Work & recreational cores

In Sec. S3.2 we determined for each gathering whether it occurred on or off campus. For cores this distinction is not possible since cores consist of multiple gatherings that in principle can occur both on or off campus. Instead we count the number of times each core appears on and off campus and perform a majority voting. Cores that have a higher frequency of on-campus meetings are denoted as *work* cores, while we call cores that appear more frequently off-campus as *recreational* cores. In case of a tie, we label the core as recreational. Figure S21a shows the voting schedule and indicates the split, while Fig. S21b shows the waiting time probability between consecutive meeting events. For work cores the waiting time probability shows clear signs of daily and weekly patterns, suggesting that these cores may be driven by the class schedule. Recreational cores on the other hand exhibit a more subtle pattern that slowly decays and is considerably higher during nighttimes—suggesting that two fundamentally different mechanisms drive the activity of the two groups. Splitting the number of cores per individual (Fig. S19) up into the two categories yields the results shown in Fig. S21c-d, where users have a broad degree of recreational cores, on average participating in 2.53, while the number of work cores is more localized with an average value of 2.74. According to Fig. S21e-f individuals spend more time in recreational cores than work cores, clearly depicting that context has great influence on the properties of meetings.

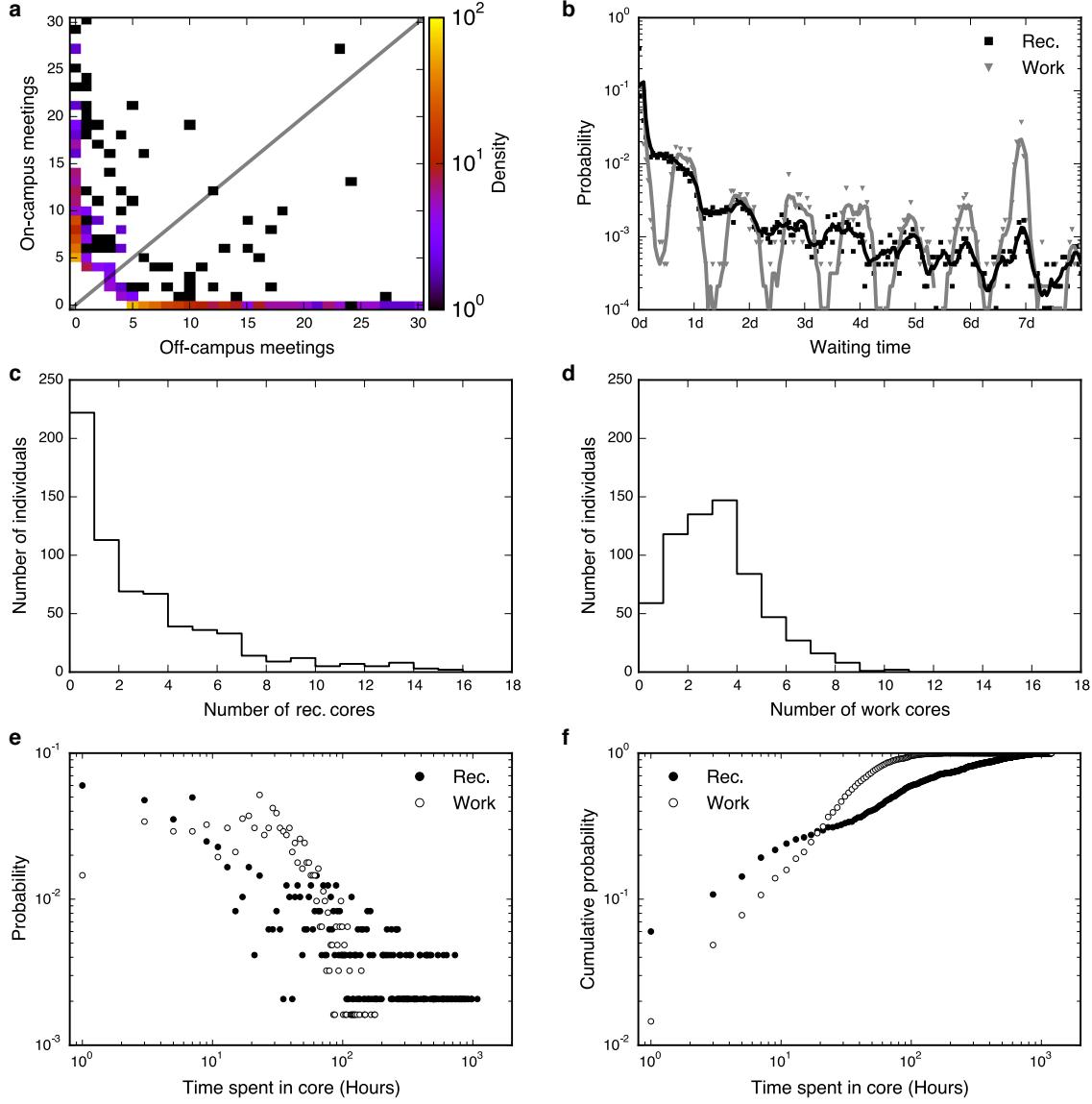


Figure S21: Distinction between work & recreational cores. **a**, Excerpt of voting scheme depicting the number of on- and off-campus meetings for each core. Gray line splits the area into work (above) and recreation (below) categories. **b**, Inter event time distributions between consecutive meetings, aggregated across all cores of similar class (work/recreation). Events are hourly binned and full lines denote moving averages, calculated using 4-hour windows. **c**, Number of work cores per user. **d**, Number of recreational cores per user. **e**, The distributions of how much time individuals, in total, spend in cores with work and recreational context. The average individual spends approximately 38 hours in a work setting, and 120 hours in a recreational context during the period of study. **f**, Cumulative probability distribution of data shown in panel e.

S4.5 Meeting regularity

Each core has a specific temporal pattern linked to it denoting the periods of time it has been present, see Fig. S22a for an example. The information contained in each pattern can be quantified using Shannon entropy [19], defined as

$$H = - \sum_t p_t \log_2 p_t, \quad (\text{S12})$$

where the sum runs over all temporal bins t and p_t is the probability of observing a specific core within given time-bin. For each core we aggregate its meeting patterns across the full study duration into weekly 1-hour bins, then within each bin we calculate the probability of observing the core. Entropy is calculated individually for cores using Eq. S12. According to Fig. S22b there is clear differences between the meeting patterns for work and recreational cores. Further, the entropy distributions (Fig. S22c) reveal that recreational cores, on average, have higher entropy and thus lower meeting regularity—indicating that they do not meet within pre-scheduled temporal-bins.

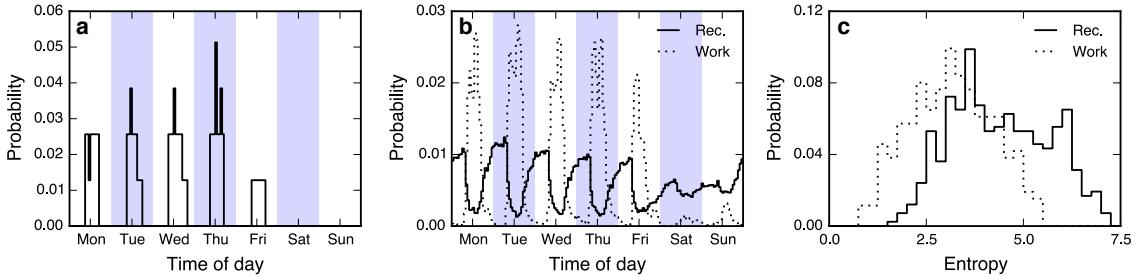


Figure S22: Meeting regularity of cores. **a**, Example pattern for a single core, denoting the probability of observing the core. Data is aggregated across all weeks into 1-hour wide bins. **b**, Aggregated meeting patterns across all cores, showing the probability of observing work and recreational cores. **c**, Distributions of meeting time entropy calculated across all cores and divided up into work and recreation categories.

S4.6 Ego viewpoint

So far we have mainly focused on the overall structural and dynamical features of cores, but we can reverse the perspective and look at cores from the perspective of individuals. From this perspective, each core provides a social context for the situation in which a person is embedded, whether it is in a work, recreation, or other relation. Figure S23a shows the involvement of a representative individual that is involved in multiple cores, producing hierarchically nested and overlapping structures. The corresponding temporal pattern of cores (Figure S23b) reveals a complex behavior.

S5 Predicting behavior from routine

In this section we describe the detailed analysis leading up to routine-based geographical location and social context prediction presented in the main text.

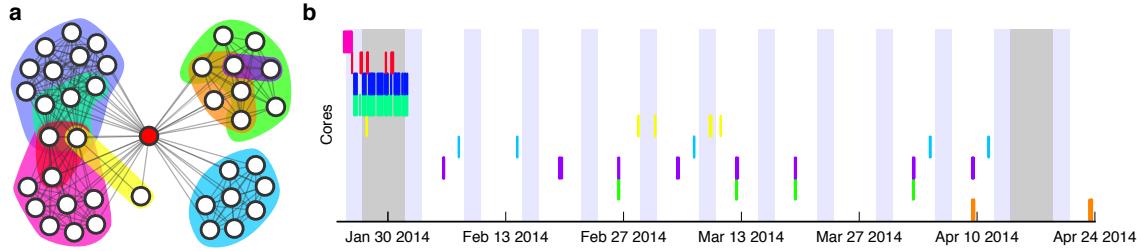


Figure S23: *Ego-centric perspective of cores for a representative individual.* **a,** Network perspective, revealing overlapping and nested structures. **b,** Temporal dynamics of individual cores, colored accordingly.

Song *et al.* have used entropy, an information theoretic measure in order to estimate the upper bound of the predictability of individuals' mobility patterns [20]. We argue here, that in analogy to geospatial behavior, human social life can be described by a temporal sequence of 'social states'. These can be used to quantify the predictability of social life.

Given a sequence of states for an individual i we can define entropy in two ways. First we can think of predictability in a temporally uncorrelated sense with entropy defined as

$$S_i^{\text{unc}} = - \sum_j^{N_i} p_j \log_2 p_j, \quad (\text{S13})$$

where p_j is the probability of observing state j and N_i is the total number of states observed by person i . Eq. S13 captures the uncertainty of your location history without taking the order of visits into account, thus discarding information contained in the daily, weekly and monthly sequences of behavior. A more sophisticated measure that includes temporal patterns is *temporal entropy*:

$$S_i^{\text{temp}} = - \sum_{T'_i \subset T_i} p(T'_i) \log_2 [p(T'_i)], \quad (\text{S14})$$

where $p(T'_i)$ is the probability of finding a subsequence T'_i in the trajectory T_i . From the entropy one can estimate the upper bound of predictability (Π_i) by applying a limiting case of Fano's inequality [20–22]:

$$S_i = H(\Pi_i) + (1 - \Pi_i) \log_2(N - 1), \quad (\text{S15})$$

where $H(\Pi_i) = -\Pi_i \log_2(\Pi_i) - (1 - \Pi_i) \log_2(1 - \Pi_i)$.

S5.1 Comparing with previous studies

In the main manuscript (Fig. 3a) we show that our ability to predict the location of individuals has an upper bound of 71% which is significantly lower than the 93% reported by Song *et al.* [20]. The main reasons behind this discrepancy are discussed below.

S5.1.1 Difference in populations

Our study population is comprised of university students that (1) not necessary have a single home/nightly location, (2) have a rich free time which they can utilize to explore new locations, and (3) have multiple work locations due to classes being distributed across a large university campus.

In contrast, Song et al. studied a sample of 50 000 individuals selected from a total population of ~ 10 million anonymous mobile phone users. The users were chosen according to (1) their travel patterns, where individuals had to visit more than two cell tower locations during the observational period of three months and (2) their average call frequency which had to be $\geq 0.5 \text{ hour}^{-1}$, effectively selecting individuals with at minimum of 12 calls per day.

S5.1.2 Geospatial resolution

Previous studies have applied call detail records (CDR) as proxy for location, inferring the position of individuals depending on which cell tower their mobile phone is connected to during a call [20, 23, 24]. While the granularity of cell tower locations in cities is around 800 meters it can be on the order of kilometers in more rural areas [25]. Figure S24 illustrates the effect of using cell tower data for positioning, as it can cluster otherwise distinct places together as one. Our location data on the other hand has a typical accuracy below 60 meters [1], enabling a more accurate spatial estimation.

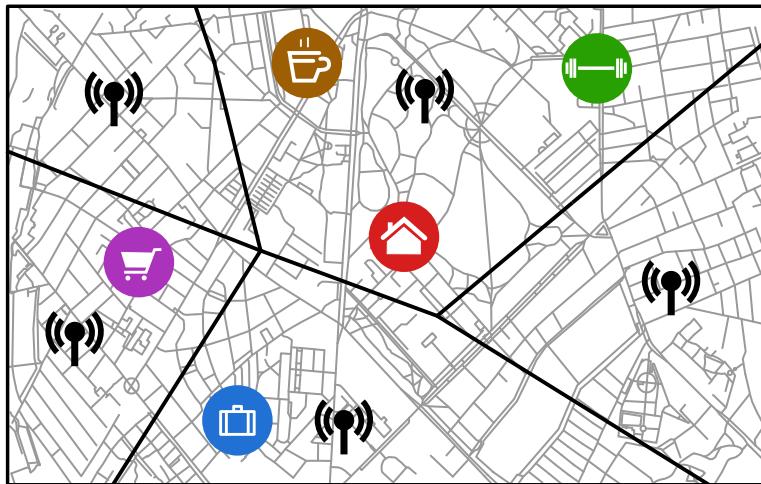


Figure S24: Cell tower resolution of a city. Otherwise distinct places can be grouped together under the same cell tower, coarse graining the geospatial position and increasing the probability of guessing correct.

S5.1.3 Effects of binning

Song et al. [20] chose, because of data granularity, to segment their data into one hour time-windows. Since our data has finer granularity, down to the minute scale, we can choose

a different resolution, but which is optimal?

We show in Fig. S25a that the finer we segment time, the better we are able to predict your location in the next time-bin. By reducing temporal bin size it is possible to achieve arbitrarily high levels predictability because segmenting data into finer time-windows increases the number of bins, which in turn leads to respective states obtaining a higher frequency of visits.

The bin-size affects both temporal and uncorrelated entropy, however, so one hypothesis is that it is still meaningful to consider the ratio between the temporally uncorrelated entropy and the temporally correlated entropy. To investigate, we study on the ratio $S_i^{\text{temp}}/S_i^{\text{unc}}$, evaluated across all individuals i . As is clear from Fig. S25b the ratio shifts towards lower values for small windows, revealing that the choice of bin width has a greater impact on temporal entropy. This suggests that predictability is greatly influenced by the choice of time-window, as has also been noted by [22, 26]. In short, considering the ratio $S_i^{\text{temp}}/S_i^{\text{unc}}$ does not solve the binning problem. Ultimately this implies that the smaller bins we apply the better we are at predicting. Because we currently are unaware of any timescale that is fundamentally descriptive of human behavior, we chose to work with temporal sequences in their natural form instead of segmenting them, predicting ‘next state’ rather than ‘state of next time bin’ (see Fig. S26).

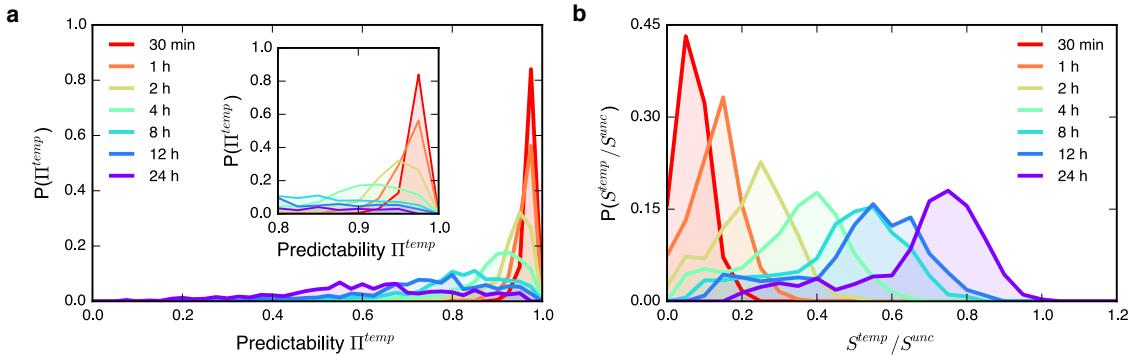


Figure S25: Effects of binning on predictability bounds. **a**, Predictability as function of window size. Inset shows a close up for high values of predictability. Segmenting time into finer bins yields higher bounds of prediction. **b**, Ratio between temporal and uncorrelated entropy values. As the size of time-windows is narrowed the ratio shifts towards zero, indicating that $S^{\text{temp}} \ll S^{\text{unc}}$.

S5.2 Data for prediction

S5.2.1 Social vocabulary

We describe the social life of an individual based on the context provided by cores, where we focus on individuals that appear in frequently observed cores (observed, on average, at least once per month). In order to include the full social life of an individual we incorporate the context provided by dyads as well as the information contained in infrequently observed cores. Note that if a core/dyad is only observed once then we denote it as a ‘noise’ state. In addition we construct one supplementary context: ‘alone’ denoting periods of time where an individual is not socially active. Following Song et al. [20] we construct a

time-series of social contexts for each user. However, as noted above, we do not segment the sequence into temporal bins, but keep it in its natural form, and predict ‘next state’ (illustrated in Fig. S26) for both geospatial and social prediction.



Figure S26: Time-series of states. Left part of the figure presents social states as they naturally occur in a temporal sequence, with Δt denoting a segmentation of the sequence within arbitrary sized time-windows. Right panel illustrates the difference in vocabularies, where the ‘natural’ sequence focuses on the order of states, while ‘segmented’ also weights states according to their duration.

S5.2.2 Location vocabulary

In addition to social context, we also collect geographical traces for each user, enabling us to reconstruct their mobility patterns. To infer context from raw location traces we use the same definitions as Cuttone *et al.* [27], where a point of interest (POI) is a location of relevance for a person, such as home, work, or a cinema. POIs are inferred by applying a density based clustering algorithm [28], with a density grouping distance of 60 meters and requiring stops to consist of at least two samples, meaning that a person must have spent a minimum of 15 minutes in the same location.

S5.2.3 Convergence of states

Figure S27 shows the distribution of the number of distinct social and geospatial states for increasing windows of time. After 90 days both probability distributions converge, implying that the number of states visited by users is saturated, indicating that we can uncover a majority of states frequented by individuals. We, however, expect this saturation only to be meta-stable, because the social networks change across adulthood [29].

S5.3 Prediction

Following Eq. S13-S15 we first calculate the respective entropies of the behavioral patterns for each individual. We show in Fig. S28a that our social patterns have lower entropy than our mobility. The figure shows that an average person approximately occupies $2^2 \approx 4$ social states and $2.5^2 \approx 6.25$ location distinct states. We also find that humans are potentially more predictable based on their social contexts than their past locations, see Fig. S28b. Previous studies have found higher levels of predictability [20, 30], see Sec. S5.1 for a full discussion.

Fig. S29a shows the interrelation between S_{social}^{unc} and $S_{location}^{unc}$, surprisingly there is no correlation between the two measures (Spearman correlation $\rho = 0.053$, p -value = 0.191), indicating that humans can be highly predictable in a social sense but very unpredictable

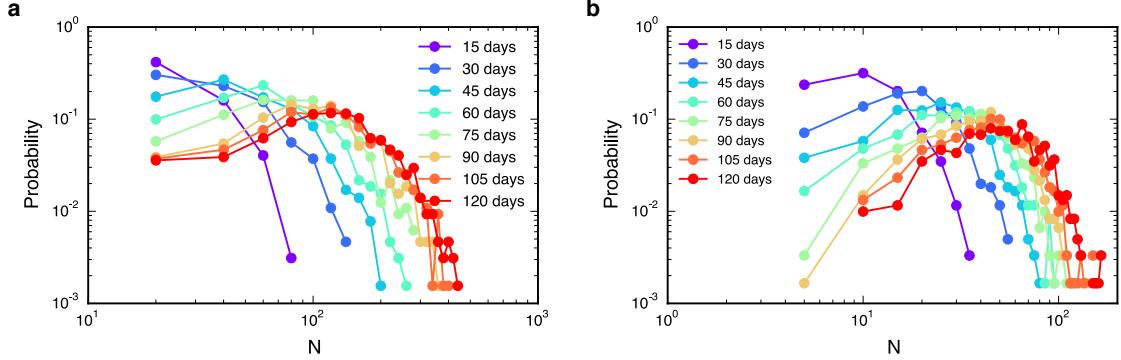


Figure S27: Distribution of the number of distinct states within a time window. **a**, Convergence of the number of social social states, showing saturation after a time window of 90 days. **b**, Saturation of visited locations, also convergence after 90 days.

location-wise and vice versa. A similar lack of correlation is observed for S^{temp} (Spearman correlation $\rho = -0.008$, p -value = 0.84), see Fig. S29b.

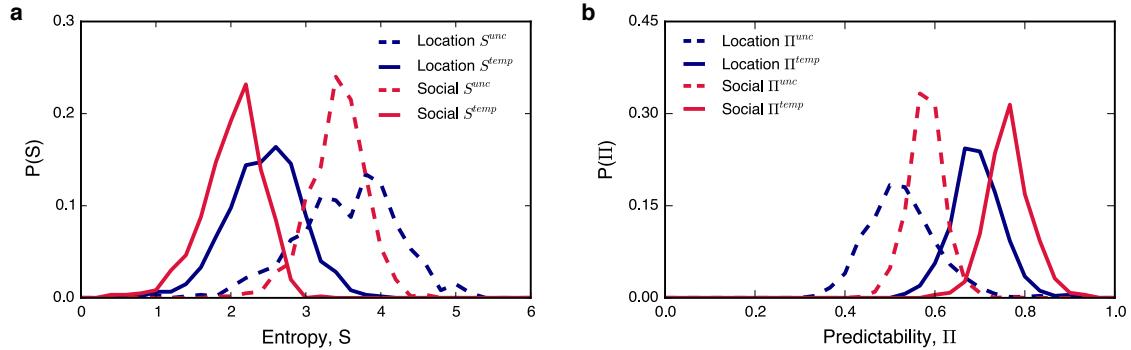


Figure S28: Probability distributions of entropy and predictability. **a**, Distributions of temporal and uncorrelated entropy for location and social behavior. As expected, temporal patterns contain more information than just frequency of visits, and hence have a lower entropy. **b**, Predictability distributions for uncorrelated and temporal patterns. On average, our social behavior is more predictable than our geospatial behavior.

S5.4 Temporal aspect of predictability

According to Fig. S29 there is no correlation between social and geospatial aspects of human life, but this is measured in terms of overall dependence. In real life we have varying degrees of predictability, a simple way of visualizing this is to look at the number of states as function of time. According to Fig. S30a we have a low number of location states during nights (resulting in low entropy) because we mainly sleep at a single location, while during days and evenings our entropy is higher, in part because we occupy more states. Note that there is very little overlap between locations visited in each 8-hour bin, e.g. morning locations are different from day locations. On Fridays we visit more locations than

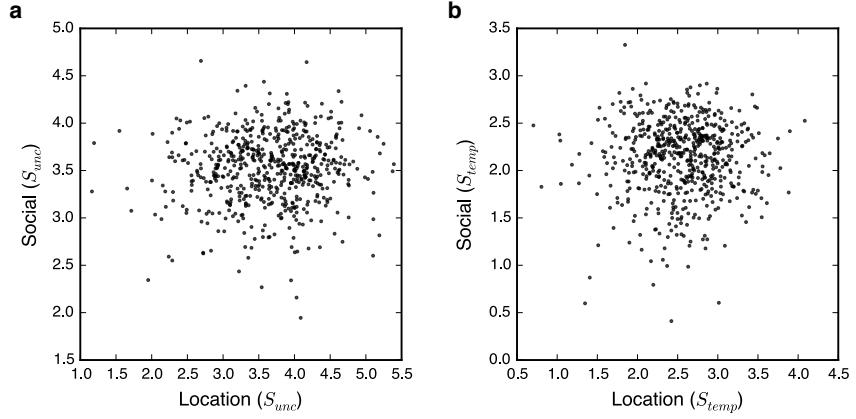


Figure S29: Correlation between social and location entropy. **a**, Mutual dependence between uncorrelated entropy for social and location states per individual. **b**, Correlation between temporal entropies for social and location states.

any other day, while during weekends we are more stationary. If we, however, consider the total number of distinct visited places (Fig. S30b) we see that Fridays and Saturdays are special because those days are used to explore new locations. Therefore, predicting location during weekends based on routine is more difficult, since we have higher entropy during these periods. Our social behavior (Fig. S30c) resembles our mobility, where we socialize mainly during the day and less during the night. Weekends are again special, interestingly we here observe a drop in in the number of social states, because we are not required to go to work or school. Fig. S30d shows that the number of social states decreases during weekends, meaning our participants reserve weekends to socialize with a few selected friends.

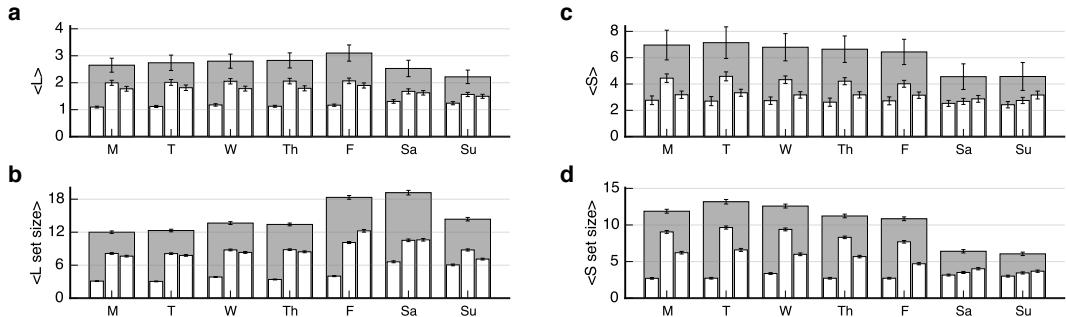


Figure S30: Nested histograms showing the temporal aspects of predictability. Binned using daily and 8-hour intervals (12 am - 8 am, 8 am - 4 pm, 4 pm - 12 am), outer bars (gray) denote days while inner bars (white) denote 8-hour windows. Bars do not necessary add up, because one can have an overlap of states between the 8-hour bins. All values are averaged across the student population. **a**, Number of average observed locations per bin. **b**, Total number of visited distinct locations. **c**, Average number of social states per time-bin. **d**, Total number of distinct social states.

Here, it is important to realize that we only observe social interactions among participants of the experiment. While the geospatial data is sampled evenly over the observation period, social interactions might have a potential bias, as it is possible to go out with non-university

friends on Fridays and weekends. We address this issue, in Fig. 5b (main manuscript), by quantifying the geo-spatial behavior of cores. Whenever a core is present we pool together the geographic locations of its individuals, bin the resulting list of (latitude, longitude)-points in 0.01×0.01 cells, and quantify the behavior using uncorrelated entropy. Using 0.1×0.1 and 0.001×0.001 cells produces comparative results. Displayed in the figure is the average entropy, averaged over all cores and binned in 8-hours bins (as described above).

S6 Social prediction

It has previously been shown that spatial behavior between individuals that share a social tie is correlated [31–34]. But the onset of co-presence lacks a temporal signature, so while the spatial traces overlap it is not generally possible to specify exactly when a friend is predictive for an individual’s behavior. Cores, however, do provide such context. An incomplete set of members provides a clue that a social interaction is about to occur (i.e. the final group member is about to arrive).

We test this concept on cores of size three; thus provided we observe two members we measure the probability of the last member arriving within the next hour. To avoid testing the hypothesis on scheduled meetings we focus on weekday nights (6 pm - 8 am) and weekends. This is the period where routine driven prediction is at its weakest. Further, in order to avoid circularity, we evaluate the hypothesis on a test month (May 2014), during which we have not identified gatherings.

S6.1 Null models

For evaluation purposes we compare cores to two reference models, both generated from real world data. We segment interactions into undirected and unweighted daily graphs, see Fig. S31a. The first null model, which we denote *random*, constructs reference groups by randomly drawing nodes from each daily graph. The second model utilizes a breadth-first search in the daily graph to create reference groups starting from a randomly chosen seed node by searching its local neighborhood. A new seed node is chosen if the search is restricted to components with fewer than three members. *BFS* is a strict null model and requires all nodes to have shared a physical interaction. We disregard reference groups if they happen to be identical to a core.

S6.2 Comparison

We test the hypothesis on a sample of 340 frequently occurring cores and 10 000 reference groups for $n = 100$ independent trials. For 33 cores we never observe an incomplete set of members during the month of May, therefore they cannot be used for prediction and are disregarded.

As shown in Fig. S31b approximately 50% of cores are predictive, in stark comparison to BFS reference groups, with the random null model performing even worse.

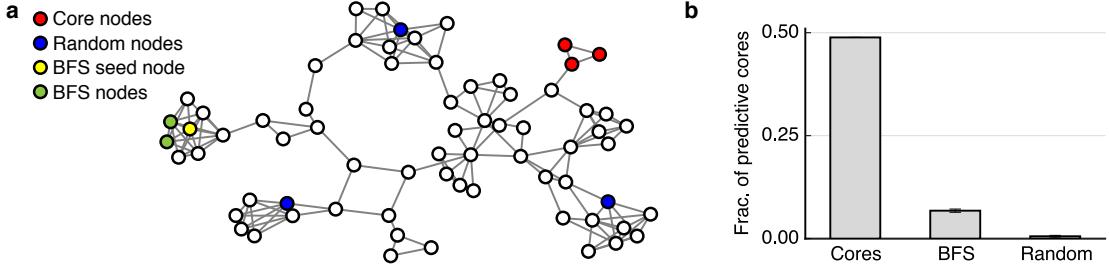


Figure S31: *Social prediction.* **a**, Daily graph of interactions, illustrating cores and construction of null models. **b** Percentage of socially predictive cores within each category. A group is predictive if it at least once has correctly predicted the arrival of the missing individual. For the reference models errorbars are calculated across $n = 100$ independent trials.

S7 A dynamic random geometric graph model

In order to investigate the statistical significance of the results found in the empirical data, we create a null model to simulate full network dynamics.

The simplest candidate for a model system is based on random geometric graphs (RGGs) [35] which incorporate random walks. We model a scenario where individuals move randomly and connect to each other when they are physically proximate. This model has three parameters. The random walker step size ℓ , population density ρ and contact radius r (of which only two are free after rescaling). In the current setting, it is natural to choose a scale for this model that lets us interpret the results in units that correspond to our actual real-world system, thus we connect two links if their distance is less than $r = 10$ meters, corresponding to the typical Bluetooth range used in the main text.

S7.1 Single time-slice RGG parameters

Random geometric graphs have been used as models for complex networks in the literature, see Ref. [36] for an overview. In a RGG, we expect to find a single giant connected component when r is much greater than the typical distance between neighboring nodes. When r is much smaller than the typical distance between neighboring nodes, the random geometric graph has many isolated components with the component size distribution approximately exponential [37]. Near the critical threshold r_c , the distribution of component sizes follows a power-law [37].

For our purposes, RGGs are interesting as a model for our real-world system around r_c because gatherings in the real-world network display a power-law distribution of sizes, as suggested by the distribution of gathering sizes displayed in Figure S9. Staying with parameters chosen to correspond to the empirical data, we construct a random geometric graph with $n = 800$ nodes distributed randomly in a $L \times L$ square, corresponding to density $\rho = n/L^2$, see Fig. S32A for an illustration, to remain consistent with the literature, we use open boundary conditions when forming components [37].

Using the interpretable ‘physical’ parameter choices described above, we can estimate ρ based on the critical threshold for RGGs reported in the literature [37, 38]. This yields $\rho = 0.014$ nodes/ m^2 , in accordance with our numerical simulations, see Fig. S32B (inset 1).

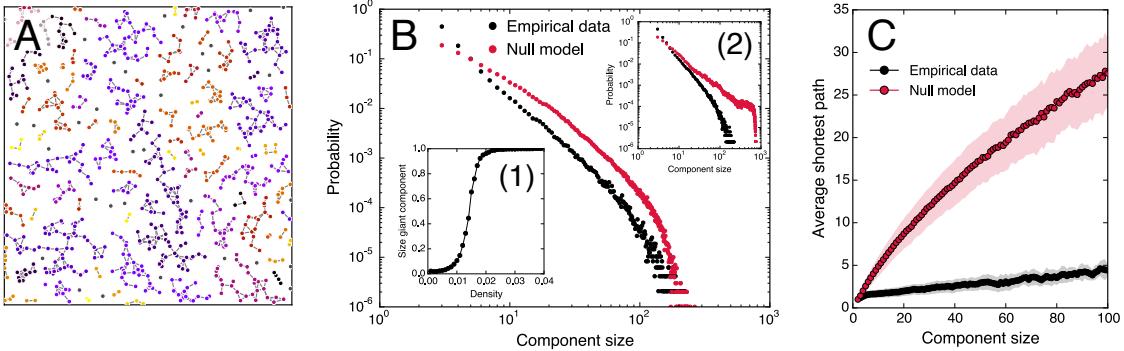


Figure S32: The RGG model (parameters as described in the main text). **A.** A snapshot of the model near the percolation threshold; notice how components do not form dense clique-like structures. **B.** The normalized distribution of component sizes in many realizations of the RGG model (red) and as observed in the empirical data (black). When the random geometric graph is near the percolation threshold, both systems display power distributions of component sizes. Inset (1) shows the average largest Connected Component size (normalized to total number of nodes) as a function of density. Inset (2) shows the real and model component size distribution at the critical point. **C.** The Average shortest path length inside components in the RGG (red). Shaded areas denote standard deviation. The Average shortest path length inside components in real data (black). Notice how the average shortest path length tends to be significantly larger in the sparse, spatially extended components generated by the RGG model.

The resulting distribution is shown in Fig. S32B (inset 2). At the critical point, the distribution of component sizes in the RGG model is broader than the distribution found in real-world data and with a noticeable cut-off due to the finite size effects. By reducing the model's density by a small amount (to $\rho \approx 0.01$), we find that the distribution of component sizes in the RGG model can be tuned to match the real-world component size distribution reasonably well. (Details on distribution matching can be found below.) The final single time-slice RGG component size distribution for the ‘physical’ parameters defined above is shown alongside its real-world counterpart in the main panel of Fig. S32B

S7.2 Model versus the real world

While the distributions of component sizes in the model is qualitatively similar to what we find in the real data, it is important for our understanding of the model’s limitations to point out that there are significant differences between the RGG and real-world data.

Firstly, the spatial distribution of node-locations is *very* different, from $\rho = 0.01$ nodes/ m^2 and $n = 800$ nodes and resulting $L \approx 283\text{ m}$. In contrast, the real data-set finds individuals distributed in small clusters throughout the entire greater Copenhagen region ($L_{\text{real}} \sim 20\,000$ meters, cf. Ref [1]).

Secondly, while both the real data and the RGGs have power law distributed component sizes, see Figure S32B, the internal structure of components in the two models are quite different. In Figure S32A we show a realization of a RGG for the final parameter values discussed above. Nodes that are within a radius of $r = 10\text{ m}$ are connected by a link. While more clustered than components in a Erdős-Rényi graph [37], the component we find in the

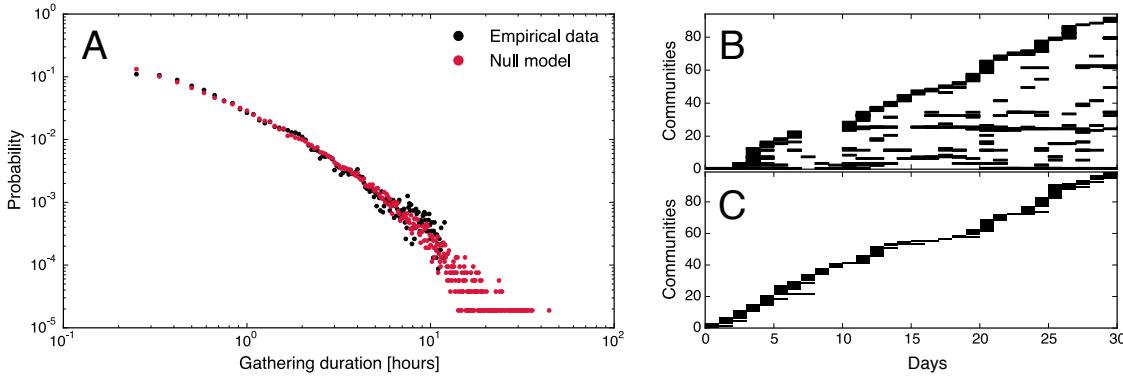


Figure S33: **A.** Distribution of gathering lifetimes for parameter values defined in the text. **B** A example of participation in cores in the real world data (from main text Fig. 3d). **C** An example of participation in cores in the dynamic RGG model. Notice how the dynamic RGG does not have recurring gatherings.

RGG model are not clique-like as we observe in the empirical data (cf. main text Figure 1). Instead of being clique-like, the RGG components have long average internal paths and significant spatial extension (the distance between extremal points often representing a significant fraction of the side-length L). One way of quantifying the differences between the empirical components and the RGG components is considering the average shortest path length inside components in the two models. This is what we explore in Figure S32C, where we quantify how paths are significantly longer in the random geometric graph than what we observe in the more clique-like components observed in the empirical data.

S7.3 A dynamic RGG model

So far we have only considered static random geometric graphs—describing how to generate a single time-slice with statistics that match our real-world data. We now implement graph dynamics by letting each node move according to a random walk; we use periodic boundary conditions for the random walkers. We implement the random walk as follows. At time step t , node i with location $(x_i(t), y_i(t))$ updates its position by selecting an angle $\theta_i(t) \in [0, 2\pi]$ at random and moving with step length ℓ in that direction, ending up in $(x_i(t+1), y_i(t+1)) = (x_i(t) + \ell \cos[\theta_i(t)], y_i(t) + \ell \sin[\theta_i(t)])$, modulo the periodic boundary conditions.

We choose step size ℓ to produce gathering lifetimes corresponding to the empirical distribution, given the choices of values for n, r, ρ listed above. We match up the two distributions by iterating over values of ℓ and pick the step size which minimizes the Kullback-Leibler (KL) divergence [39] between the two probability distributions. We find that $\ell = 0.65$ (i.e. $0.65m$ in the physical interpretation) produces lifetimes with a distribution of gathering lifetimes similar to that of the empirical system, see Figure S33A for the two distributions.

Thus, quite remarkably, the very simple dynamic RGG model based on random walks for each node is able to reproduce, not only the the distribution of gathering sizes (Figure S32B),

but also the distribution of gathering lifetimes. This implies that in terms of *gatherings* this very simple model is able to produce behavior quite similar to what we observe in the empirical data—at least in terms of distributions.

S7.4 Random walks in 2D

Let us quickly recapitulate the central results on random walks in 2D*. The first key result is that the average displacement of a random walker is zero. This result simply means that if we choose a point (x_0, y_0) and start an ensemble of random walkers there, each one moving according to the rules outlined above, their *average* location will remain (x_0, y_0) independent of the number of steps. The second result is that after T steps, the root mean square displacement d_{rms} of a walker with stepsize ℓ is $d_{\text{rms}} = \ell\sqrt{T}$.

Taken together these two results characterize the behavior of individual random walkers. A walker does not remain at its starting point, but tends to drift, on average traveling some distance d in $T \approx (d/\ell)^2$ steps. The fact that the average displacement is zero simply implies that a walker is equally likely to drift in any direction. Thus if we find multiple walkers close to each other—as is the case for a gathering in our model—we expect the gathering to disappear after some time T , as the members each drift in a random direction.

S7.5 The dynamic RGG model does not generate recurring meetings

So far, we have shown that the dynamic RGG model is able to reproduce the distributional properties of our real-world data. That is, for a single time-slice, the model can generate component size distributions that are quite similar to what we observe in the empirical data. When we add a dynamic component to the model, allowing each node to move according to a random walk, we see that the real-world distribution of gathering life-times can also be matched by this simple model.

The dynamic RGG model, however, does not generate cores[†]. Recall that cores are *recurring* gatherings, where a specific group of individuals meets repeatedly over time. Given the known dynamics of random walkers summarized above, the long-term picture finds each random walker drifting according to its own trajectory. Since there are no correlations between individual trajectories, the probability that the members of any gathering with more than two nodes will be in close proximity again, on the time-scale we are studying here (e.g. simulating \sim one month of data with parameters as above), is vanishing. Thus, we do not expect to see cores occurring more than once in the dynamic RGG model.

This is precisely what we observe when we analyze the model data. Figure S33B is a similar to the main text’s Figure 3d, but this time showing cores with any number of occurrences (not just the recurring gatherings), displaying how a representative individual participates in her/his cores as a function of time. Time runs along the x -axis and each vertical line represents an activation of that core. In the main text, we call this the path

*Random walks in 2D are well understood, see e.g. <http://mathworld.wolfram.com/RandomWalk2-Dimensional.html>/ for a nice overview

[†]Technically, a given group of people only meeting a single time is both a gathering and a core (a core which occurs only once). Here, we use core in the sense that we do in the main paper—to signify recurring gatherings consist of 3 or more individuals and meet more than once per month on average

through the cores an individual’s ‘social trajectory’. In Figure S33B, we have generated a social trajectory for a representative node in the dynamic RGG model. The differences in core participation patterns illustrates the argument above nicely. Over the same time-period, the dynamic RGG model-node in question participates in nearly a similar number of social contexts, but all of them occurring only once; sometimes gatherings in the model can last multiple days. In the real-world data, however, a pattern of recurring meetings is evident. It is based on these patterns that we use the cores to make predictions.

Thus, it is clear that due to the lack of recurring cores, the dynamic RGG model does not have predictability in the senses that we discuss in the main text. Let us be more explicit about this point. Recall that in the main text, we define predictability in two senses. Below, we discuss each one in detail.

- Firstly, we discuss what we call the ‘social unit property’. The basic idea behind the social unit property is that, cores represent meetings between $n > 2$ individuals, characterized by the fact that all n members are usually present when the core is active. Therefore an observation of an incomplete set of core members implies that the remaining members will arrive shortly, a fact which can be used for prediction. It is trivial to see that cores in the dynamic RGG model do not possess the social unit property, simply because cores typically only occur once.
- Secondly, we measure predictability as defined in Ref. [20] (see main text for details). The calculation for predictability in this sense depends on compressing the social trajectory. As is clear from Figure S33B, the social trajectory for a node in the dynamic RGG model simply grows over time, continually introducing new gatherings that only occur once. This implies that the entropy for a node continues to grow as a function of time. The fact that the entropy keeps growing should not surprise us, given that social structures arising in the simple RGG model are based on a random process. The fact that the entropy is unbounded implies that there is zero predictability in the Ref. [20] sense as well.

As noted earlier, RGG’s have been used as models for social systems in the literature. Specifically, Refs. [40, 41] have explored modeling social networks using variations of the simple RGG model explored above, also finding that distributional properties of empirical data can be matched quite well. In both cases the walkers ‘slow down’ (decrease step size) upon forming links with other agents. This social slow-down results in denser (more realistic) gatherings, but due to the fundamentally random nature of walkers in each model, neither of these model display cores that appear multiple times. Thus, just like the simple dynamic RGG, these models do a surprisingly good job of simulating behavioral data for a single day (creates gatherings), but none of them capture behavior across weeks and months (unable to create realistic cores).

S7.6 Summary

In summary, the random geometric graph with random walkers provides an interesting model which is able to reproduces distributional properties of component sizes as well as component life-times quite accurately. Therefore this model is an excellent point of

comparison, when we want to understand the significance of our findings regarding the empirical dynamic networks.

There are some important differences between the model and the empirical data. When we look at the RGG components, for example, we find that their spatial configuration is quite different from what we observe in real data. In order to reproduce the distributional properties, all model components must be packed in a very small area (a square with side length $L \approx 280$ meters) and the model component are sparser with much longer path-lengths than components in the real-world data.

The most important difference between the empirical data and the simple RGG model, however, is that the latter does not display any meaningful core structure. As we have shown above, this implies that the predictability we observe in the empirical dataset is not replicated by the dynamic RGG model and can be considered a real effect expressed in the data.

S8 Coordination of meetings

This section describes how we calculate the amount of coordination leading up to gatherings. Prior to a meeting individuals might need to coordinate about when and where to meet. This coordination can be conveyed through various means: (1) individuals can organize in real time through electronic means, such as online social networks and mobile phones, (2) they can verbally have scheduled meetings beforehand, i.e. at previous gatherings, (3) or attend routine driven pre-scheduled meetings, arranged by an institution, e.g. the university. We consider the coordination in the hours leading up to a meeting, measured in terms of increased calling and texting activity. Because calling frequencies change over the course of a day, and because individuals can have fundamentally distinct calling patterns [1, 42, 43], we compare activity leading up to a meeting to hour-by-hour dynamic individual baselines, defined as

$$c_t = \frac{1}{N} \sum_{n=1}^N \frac{a_t^n}{\bar{a}_t^n}, \quad (\text{S16})$$

where N is the number of individuals participating in the gathering, a_t^n is the activity of person n , and \bar{a}_t^n is the baseline activity. The equation denotes increased coordination levels t hours before a meeting. Because gatherings have a broad distribution of lifetimes (Fig. S9), we restrict the calculation to individuals that participate in the first hour of the meeting. According to Fig. S34a-b, meetings during the weekend require more coordination than meeting during weekdays. This implies that weekend behavior is less scheduled, emphasizing the problem of predicting behavior using traditional routine-based measures. In addition, Fig. S34c reveals that meetings, independently of size, require the same amount of coordination per person, illustrating the validity of Fig. S34a for varying gathering sizes.

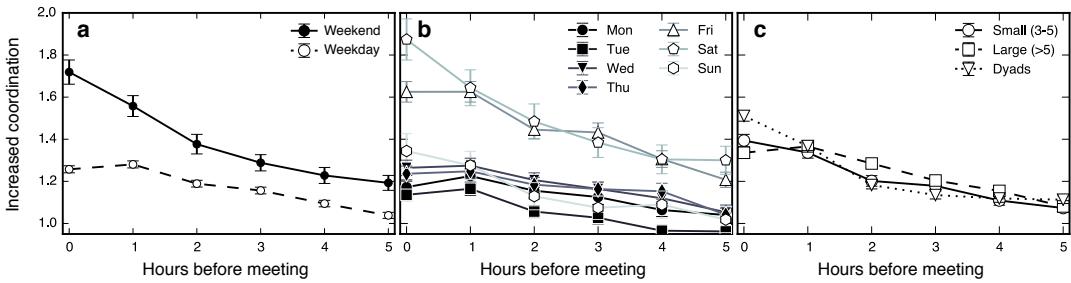


Figure S34: Increased coordination prior to a meeting. Calculated for all nodes participating in the first hour of a gathering. **a**, Required amounts of coordination between nodes, depending on when the gatherings meets. More coordination is required to organize meetings during weekends (Friday 4 pm - Sunday) than during weekdays (Monday - Friday 4 pm). **b**, Further sub-dividing the categories from panel **a** reveals that Fridays and Saturdays are special. **c**, Effect of size of the meeting on coordination. On average it requires equal amounts of coordination, per person, to organize a meeting, independent of the size of the group.

References

- [1] Stopczynski, A. *et al.* Measuring large-scale social networks with high resolution. *PLoS ONE* **9**, e95978 (2014).
- [2] Sekara, V. & Lehmann, S. The strength of friendship ties in proximity sensor data. *PLoS One* **9**, e100915 (2014).
- [3] Eagle, N., Pentland, A. S. & Lazer, D. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences* **106**, 15274–15278 (2009).
- [4] Clauset, A. & Eagle, N. Persistence and periodicity in a dynamic proximity network. *arXiv preprint arXiv:1211.7343* (2012).
- [5] Sulo, R., Berger-Wolf, T. & Grossman, R. Meaningful selection of temporal resolution for dynamic networks. In *Proceedings of the Eighth Workshop on Mining and Learning with Graphs*, 127–136 (ACM, 2010).
- [6] Krings, G., Karsai, M., Bernhardsson, S., Blondel, V. D. & Saramäki, J. Effects of time window size and placement on the structure of an aggregated communication network. *EPJ Data Science* **1**, 1–16 (2012).
- [7] Simmel, G. Quantitative aspects of the group. *The Sociology of Georg Simmel* 87–177 (1950).
- [8] Coser, L. A. & Merton, R. K. *Masters of sociological thought: Ideas in historical and social context* (Harcourt Brace Jovanovich New York, 1971).
- [9] Ward Jr, J. H. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association* **58**, 236–244 (1963).

- [10] Gower, J. C. & Ross, G. Minimum spanning trees and single linkage cluster analysis. *Applied statistics* **54**–64 (1969).
- [11] Greene, D., Doyle, D. & Cunningham, P. Tracking the evolution of communities in dynamic social networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on*, 176–183 (2010).
- [12] Newman, M. E. & Girvan, M. Finding and evaluating community structure in networks. *Physical review E* **69**, 026113 (2004).
- [13] Ahn, Y.-Y., Bagrow, J. P. & Lehmann, S. Link communities reveal multiscale complexity in networks. *Nature* **466**, 761–764 (2010).
- [14] Palla, G., Barabási, A.-L. & Vicsek, T. Quantifying social group evolution. *Nature* **446**, 664–667 (2007).
- [15] Reis, H. T. & Wheeler, L. Studying social interaction with the rochester interaction record. *Advances in experimental social psychology* **24**, 269–318 (1991).
- [16] Davis, A., Gardner, B. B. & Gardner, M. R. *Deep South; a social anthropological study of caste and class.* (University of Chicago Press, 1941).
- [17] Tibshirani, R., Walther, G. & Hastie, T. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**, 411–423 (2001).
- [18] Milligan, G. W. & Cooper, M. C. An examination of procedures for determining the number of clusters in a data set. *Psychometrika* **50**, 159–179 (1985).
- [19] Shannon, C. A mathematical theory of communication. *The Bell System Technical Journal* **27**, 379–423 (1948).
- [20] Song, C., Qu, Z., Blumm, N. & Barabási, A.-L. Limits of predictability in human mobility. *Science* **327**, 1018–1021 (2010).
- [21] Fano, R. M. *Transmission of information: a statistical theory of communications* (MIT Press, 1961).
- [22] Jensen, B. S., Larsen, J. E., Jensen, K., Larsen, J. & Hansen, L. K. Estimating human predictability from mobile sensor data. In *Machine Learning for Signal Processing (MLSP), 2010 IEEE International Workshop on*, 196–201 (IEEE, 2010).
- [23] Bagrow, J. P. & Lin, Y.-R. Mesoscopic structure and social aspects of human mobility. *PloS one* **7**, e37676 (2012).
- [24] Gonzalez, M. C., Hidalgo, C. A. & Barabasi, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
- [25] Trevisani, E. & Vitaletti, A. Cell-id location technique, limits and benefits: an experimental study. In *Mobile Computing Systems and Applications, 2004. WMCSA 2004. Sixth IEEE Workshop on*, 51–60 (IEEE, 2004).

- [26] Lin, M., Hsu, W.-J. & Lee, Z. Q. Predictability of individuals' mobility with high-resolution positioning data. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 381–390 (ACM, 2012).
- [27] Cuttone, A., Lehmann, S. & Larsen, J. E. Inferring human mobility from sparse low accuracy mobile sensing data. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, 995–1004 (ACM, 2014).
- [28] Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, vol. 96, 226–231 (1996).
- [29] Wrzus, C., Hänel, M., Wagner, J. & Neyer, F. J. Social network changes and life events across the life span: A meta-analysis. *Psychological bulletin* **139**, 53 (2013).
- [30] Lu, X., Bengtsson, L. & Holme, P. Predictability of population displacement after the 2010 haiti earthquake. *Proceedings of the National Academy of Sciences* **109**, 11576–11581 (2012).
- [31] Crandall, D. J. *et al.* Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences* **107**, 22436–22441 (2010).
- [32] Cho, E., Myers, S. A. & Leskovec, J. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1082–1090 (ACM, 2011).
- [33] Wang, D., Pedreschi, D., Song, C., Giannotti, F. & Barabasi, A.-L. Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1100–1108 (ACM, 2011).
- [34] De Domenico, M., Lima, A. & Musolesi, M. Interdependence and predictability of human mobility and social interactions. *Pervasive and Mobile Computing* **9**, 798–807 (2013).
- [35] Penrose, M. *Random geometric graphs* (Oxford University Press Oxford, 2003).
- [36] Barthelemy, M. Spatial networks. *Physics Reports* **499**, 1 – 101 (2011).
- [37] Dall, J. & Christensen, M. Random geometric graphs. *Phys. Rev. E* **66**, 016121 (2002). URL <http://link.aps.org/doi/10.1103/PhysRevE.66.016121>.
- [38] Quintanilla, J., Torquato, S. & Ziff, R. M. Efficient measurement of the percolation threshold for fully penetrable discs. *Journal of Physics A: Mathematical and General* **33**, L399 (2000). URL <http://stacks.iop.org/0305-4470/33/i=42/a=104>.
- [39] Kullback, S. & Leibler, R. A. On information and sufficiency. *The annals of mathematical statistics* **22**, 79–86 (1951).

- [40] González, M. C., Lind, P. G. & Herrmann, H. J. System of mobile agents to model social networks. *Phys. Rev. Lett.* **96**, 088702 (2006). URL <http://link.aps.org/doi/10.1103/PhysRevLett.96.088702>.
- [41] Starnini, M., Baronchelli, A. & Pastor-Satorras, R. Modeling human dynamics of face-to-face interaction networks. *Physical review letters* **110**, 168701 (2013).
- [42] Onnela, J.-P. *et al.* Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences* **104**, 7332–7336 (2007).
- [43] Jo, H.-H., Karsai, M., Kertész, J. & Kaski, K. Circadian pattern and burstiness in mobile phone communication. *New Journal of Physics* **14**, 013055 (2012).