scientific reports

Check for updates

OPEN

Hyperbolic mapping of human proximity networks

Marco A. Rodríguez-Flores & Fragkiskos Papadopoulos✉

Human proximity networks are temporal networks representing the close-range proximity among humans in a physical space. They have been extensively studied in the past 15 years as they are critical for understanding the spreading of diseases and information among humans. Here we address the problem of mapping human proximity networks into hyperbolic spaces. Each snapshot of these networks is often very sparse, consisting of a small number of interacting (i.e., non-zero degree) nodes. Yet, we show that the time-aggregated representation of such systems over sufficiently large periods can be meaningfully embedded into the hyperbolic space, using methods developed for traditional (non-mobile) complex networks. We justify this compatibility theoretically and validate it experimentally. We produce hyperbolic maps of six different real systems, and show that the maps can be used to identify communities, facilitate efficient greedy routing on the temporal network, and predict future links with significant precision. Further, we show that epidemic arrival times are positively correlated with the hyperbolic distance from the infection sources in the maps. Thus, hyperbolic embedding could also provide a new perspective for understanding and predicting the behavior of epidemic spreading in human proximity systems.

Understanding the time-varying proximity patterns among humans in a physical space is crucial for better understanding the transmission of airborne diseases, the efficiency of information dissemination, social behavior, and influence[1–8]. To this end, human proximity networks have been captured in different environments over days, weeks or months[2,4,5,9–13]. Such time-varying networks are represented as a series of static graph snapshots. Each snapshot corresponds to an observation interval or time slot, which typically spans a few seconds to several minutes depending on the devices used to collect the data. The nodes in each snapshot are people and an edge between two nodes signifies that they had been within proximity range during the corresponding slot. At the finest resolution, each slot spans 20 s and the proximity range is 1.5 m. Such networks have been captured by the SocioPatterns collaboration[14] in closed settings, such as hospitals, schools, scientific conferences and workplaces, and correspond to face-to-face interactions[9–13]. At a coarser resolution, each snapshot spans several minutes and proximity range can be up to 10 m or more. Such networks have been captured in university dormitories, residential communities and university campuses[4,5,15].

Irrespective of the context, measurement period and measurement method, different human proximity networks have been shown to exhibit similar structural and dynamical properties[6,16]. Examples of such properties include the broad distributions of contact and intercontact durations[1–3,16], and the repeated formation of groups that consist of the same people[17,18]. Interestingly, these and other properties of human proximity systems can be well reproduced by simple models of mobile interacting agents[18,19]. Specifically, in the recently developed force-directed motion model[18] similarities among agents act as forces that direct the agents' motion toward other agents in the physical space and determine the duration of their interactions. The probability that two nodes are connected in a snapshot generated by the model resembles the connection probability in the popular $\mathbb{S}^1$ model of traditional (non-mobile) complex networks, which is equivalent to random hyperbolic graphs[20–22]. Based on this observation, the dynamic-$\mathbb{S}^1$ model has been recently suggested as a minimal latent-space model for human proximity networks[22]. The model forgoes the motion component and assumes that each network snapshot is a realization of the $\mathbb{S}^1$ model. The dynamic-$\mathbb{S}^1$ reproduces many of the observed characteristics of human proximity networks, while being mathematically tractable. Several of the model's properties have been proven in Ref.[22].

Our approach to map human proximity networks into hyperbolic spaces is founded on the dynamic-$\mathbb{S}^1$ model. Specifically, given that the dynamic-$\mathbb{S}^1$ can generate synthetic temporal networks that resemble human proximity networks across a wide range of structural and dynamical characteristics, can we reverse the synthesis and map (embed) human proximity networks into the hyperbolic space, in a way congruent with the model? Would the results of such mapping be meaningful? And could the obtained maps facilitate applications, such as community detection, routing on the temporal network, prediction of future links, and prediction of epidemic arrival times?

Department of Electrical Engineering, Computer Engineering and Informatics, Cyprus University of Technology, Limassol 3036, Cyprus. ✉email: f.papadopoulos@cut.ac.cy

Network	Days	N	τ	$\bar{n}$	$\bar{k}$	$\bar{\bar{k}}$	T
Hospital	5	75	17,376	2.9	0.05	30	0.84
Primary school	2	242	5846	30	0.18	69	0.72
Conference	2.5	113	10,618	3.3	0.03	39	0.85
High school	5	327	18,179	17	0.06	36	0.61
Office building	10	217	49,678	2.8	0.01	39	0.74
Friends and family	31	112	7317	58	1.5	57	0.48

Table 1. Overview of the considered real networks. N is the number of nodes, τ is the total number of time slots (snapshots), $\bar{n}$ is the average number of interacting (i.e., non-zero degree) nodes per snapshot, $\bar{k}$ is the average node degree per snapshot, $\bar{\bar{k}}$ is the average degree in the time-aggregated network formed over the full observation duration τ, and parameter T is the network temperature used in the dynamic-$\mathbb{S}^1$ model to generate synthetic counterparts of the real systems (see "Methods"). The table also shows the number of observation days for each network.

Here we provide the affirmative answers to these questions. Our approach is based on embedding the time-aggregated network of human proximity systems over an adequately large observation period, using methods developed for traditional complex networks that are based on the $\mathbb{S}^1$ model[23]. In the time-aggregated network, two nodes are connected if they are connected in at least one network snapshot during the observation period. We justify this approach theoretically by showing that the connection probability in the time-aggregated network in the dynamic-$\mathbb{S}^1$ model resembles the connection probability in the $\mathbb{S}^1$ model, and explicitly validate it in synthetic networks. Following this approach, we produce hyperbolic maps of six different real systems, and show that the obtained maps are meaningful: they can identify actual node communities, they can facilitate efficient greedy routing on the temporal network, and they can predict future links with significant precision. Further, we show that epidemic arrival times in the temporal network are positively correlated with the hyperbolic distance from the infection sources in the maps.

Results

Data. We consider the following face-to-face interaction networks from SocioPatterns[14]. (i) A hospital ward in Lyon[11], which corresponds to interactions involving patients and healthcare workers during five observation days. (ii) A primary school in Lyon[10], which corresponds to interactions involving children and teachers of ten different classes during two days. (iii) A scientific conference in Turin[9], which corresponds to interactions among conference attendees during two and a half days. (iv) A high school in Marseilles[12], which corresponds to interactions among students of nine different classes during five days. And (v) an office building in Saint Maurice[24], which corresponds to interactions among employees of 12 different departments during ten days. Each snapshot of these networks corresponds to an observation interval (time slot) of 20 s, while proximity was recorded if participants were within 1.5 m in front of each other.

We also consider the Friends & Family Bluetooth-based proximity network[5]. This network corresponds to the proximities among residents of a community adjacent to a major research university in the US during several observation months. We consider the data recorded in March 2011. Each snapshot corresponds to an observation interval of 5 min, while proximity was recorded if participants were within a radius of 10 m from each other. Thus proximity in this network does not imply face-to-face interaction. Table 1 gives an overview of the data.

$\mathbb{S}^1$ and dynamic-$\mathbb{S}^1$ models. We first provide an overview of the $\mathbb{S}^1$ and dynamic-$\mathbb{S}^1$ models. In the next section we show that the connection probability in the time-aggregated network in the latter resembles the connection probability in the former. Based on this equivalence, we then map the time-aggregated networks of the considered real data to the hyperbolic space using a recently developed method that is based on the $\mathbb{S}^1$ model.

$\mathbb{S}^1$ model. The $\mathbb{S}^1$ model[20,21] can generate synthetic network snapshots that possess many of the common structural properties of real networks, including heterogeneous or homogeneous degree distributions, strong clustering, and the small-world property. In the model, each node has latent (or hidden) variables κ, θ. The latent variable κ is proportional to the node's expected degree in the resulting network. The latent variable θ is the angular similarity coordinate of the node on a circle of radius $R = N/2\pi$, where N is the total number of nodes. To construct a network with the model that has size N, average node degree $\bar{k}$, and temperature $T \in (0, 1)$, we perform the following steps. First, for each node $i = 1, 2, \ldots, N$, we sample its angular coordinate θ_i uniformly at random from $[0, 2\pi]$, and its degree variable κ_i from a probability density function $\rho(\kappa)$. Then, we connect every pair of nodes i, j with the Fermi-Dirac connection probability

$$p(\chi_{ij}) = \frac{1}{1 + \chi_{ij}^{1/T}}.$$

(1)

In the last expression, χ_{ij} is the effective distance between nodes i and j,

$$\chi_{ij} = \frac{R\Delta\theta_{ij}}{\mu\kappa_i\kappa_j}, \tag{2}$$

where $\Delta\theta_{ij} = \pi - |\pi - |\theta_i - \theta_j||$ is the similarity distance between nodes i and j. Parameter μ in (2) is derived from the condition that the expected degree in the network is indeed $\bar{k}$, yielding

$$\mu = \frac{\bar{k}\sin(T\pi)}{2\bar{\kappa}^2 T\pi}, \tag{3}$$

where $\bar{\kappa} = \int \kappa\rho(\kappa)d\kappa$.

The degree distribution $P(k)$ in the resulting network has a similar functional form as $\rho(\kappa)$. Thus, the model can generate networks with *any* degree distribution depending on $\rho(\kappa)$. For instance, a power law degree distribution with exponent $\gamma > 2$ is obtained if $\rho(\kappa) \propto \kappa^{-\gamma}$, while a Poisson degree distribution with mean $\bar{k}$ is obtained if $\rho(\kappa) = \delta(\kappa - \bar{k})$, where $\delta(x)$ is the Dirac delta function[20,25]. Smaller values of the temperature T favor connections at smaller effective distances and increase the average clustering in the network[21]. The $\mathbb{S}^1$ model is equivalent to random hyperbolic graphs, i.e., to the hyperbolic $\mathbb{H}^2$ model[21], after transforming the degree variables κ_i to radial coordinates r_i via

$$r_i = \hat{R} - 2\ln\frac{\kappa_i}{\kappa_0}, \tag{4}$$

where κ_0 is the smallest κ_i and $\hat{R} = 2\ln[N/(\pi\mu\kappa_0^2)]$ is the radius of the hyperbolic disk where all nodes reside. After this change of variables, the effective distance in (2) becomes $\chi_{ij} = e^{\frac{1}{2}(x_{ij}-\hat{R})}$, where $x_{ij} = r_i + r_j + 2\ln(\Delta\theta_{ij}/2)$ is approximately the hyperbolic distance between nodes i and j[21]. Therefore, we can refer to the degree variables κ_i as "coordinates" and use terms *effective distance* and *hyperbolic distance* interchangeably.

Given the ability of the $\mathbb{S}^1/\mathbb{H}^2$ model to construct synthetic networks that resemble real networks, several methods have been developed to map real networks into the hyperbolic plane, i.e., to infer the nodes' latent coordinates r (or κ) and θ, according to the model[23,26–30]. The hyperbolic maps produced by these methods have been shown to be meaningful, and have been efficiently used in applications such as community detection, greedy routing and link prediction[26–35]. Model-free mapping methods have also been developed[36]. Further, on a related note, there is a large body of work on embedding both static and temporal networks into Euclidean spaces, e.g., see Refs.[37–39], and references therein. However, no prior work has considered embedding temporal networks into hyperbolic spaces, which provide a more accurate reflection of the geometry of real networks[31].

Dynamic-$\mathbb{S}^1$ model. The dynamic-$\mathbb{S}^1$ model is based on the $\mathbb{S}^1$ model and has been shown to reproduce many of the observed structural and dynamical properties of human proximity networks[22]. The dynamic-$\mathbb{S}^1$ models a sequence of network snapshots, $G_t, t = 1, \ldots, \tau$. Each snapshot is a realization of the $\mathbb{S}^1$ model. Therefore, there are N nodes that are assigned latent coordinates κ, θ as in the $\mathbb{S}^1$ model, which remain fixed. The temperature $T \in (0, 1)$ is also fixed, while each snapshot G_t is allowed to have a different average degree $\bar{k}_t, t = 1, \ldots, \tau$. The snapshots are generated according to the following simple rules:

(1) at each time step $t = 1, \ldots, \tau$, snapshot G_t starts with N disconnected nodes, while $\bar{k}$ in (3) is set equal to $\bar{k}_t$;
(2) each pair of nodes i, j connects with probability given by (1);
(3) at time $t + 1$, all the edges in snapshot G_t are deleted and the process starts over again to generate snapshot G_{t+1}.

As shown in Ref.[22], temperature T plays a central role in network dynamics in the model, dictating the distributions of contact and intercontact durations, the time-aggregated node degrees, and the formation of unique and recurrent components. Specifically, the contact and intercontact distributions are power laws with exponents $2 + T$ and $2 - T$, respectively. These exponents lie within the ranges observed in real systems[22]. Further, larger values of T increase the connection probability at larger distances, which increases the time-aggregated node degrees. For the same reason, larger values of T increase the number of unique components formed, while decreasing the number of recurrent components. See Ref.[22] for further details.

Hyperbolic mapping of human proximity networks.
Theoretical considerations. Assuming that a sequence of network snapshots $G_t, t = 1, \ldots, \tau$, has been generated by the dynamic-$\mathbb{S}^1$ model, we show below that we can accurately infer the nodes' latent coordinates κ, θ from the time-aggregated network, using existing methods that are based on the $\mathbb{S}^1$ model. This is justified by the fact that the connection probability in the time-aggregated network of the dynamic-$\mathbb{S}^1$ resembles the connection probability in the $\mathbb{S}^1$. Indeed, in the time-aggregated network two nodes are connected if they are connected in at least one of the snapshots. Assuming for simplicity that each snapshot has the same average degree $\bar{k}_t = \bar{k}$, the connection probability in the time-aggregated network of the dynamic-$\mathbb{S}^1$, is

$$P(\chi_{ij}) = 1 - [1 - p(\chi_{ij})]^\tau, \tag{5}$$

where $p(\chi_{ij})$ is given by (1). Further, as shown in Ref.[22], the expected degree of a node in the time-aggregated network, $\tilde{\kappa}$, is related to the node's latent degree κ, via
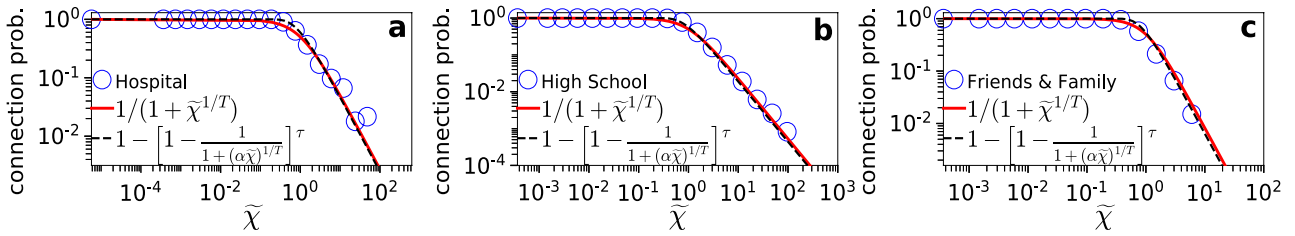
Figure 1. Connection probability in the time-aggregated network versus Fermi-Dirac connection probability. The results correspond to the synthetic counterparts of the hospital, high school and Friends & Family, constructed using the dynamic-$\mathbb{S}^1$ model as described in "Methods". The blue circles show the empirical connection probabilities. The solid red and dashed black lines correspond to (1) and (7), respectively. The values of parameters T and τ in each case are as shown in Table 1, while $\alpha = \tau^T / \Gamma(1+T)$. Similar results hold for the counterparts of the rest of the real systems (see Supplementary Information, Sect. I).

$$\tilde{\kappa} = \alpha\kappa, \tag{6}$$

where $\alpha = \tau^T / \Gamma(1+T)$ for $\tau \gg 1$, and Γ is the gamma function. Equation (6) is derived in the thermodynamic limit ($N \to \infty$), where there are no cutoffs imposed to node degrees by the network size. We can therefore rewrite (5) as

$$P(\tilde{\chi}_{ij}) = 1 - \left[1 - p(\alpha\tilde{\chi}_{ij})\right]^\tau$$
$$= 1 - \left\{ 1 + \frac{1}{\tau}\left[\frac{\Gamma(1+T)}{\tilde{\chi}_{ij}}\right]^{1/T} \right\}^{-\tau} \tag{7}$$

$$\approx 1 - e^{-\left[\frac{\Gamma(1+T)}{\tilde{\chi}_{ij}}\right]^{1/T}}, \tag{8}$$

where

$$\tilde{\chi}_{ij} = \frac{R\Delta\theta_{ij}}{\tilde{\mu}\tilde{\kappa}_i\tilde{\kappa}_j} = \frac{\chi_{ij}}{\alpha} \tag{9}$$

is the effective distance between nodes i and j in the time-aggregated network, while $\tilde{\mu} = \mu/\alpha$. The exponential approximation in (8) holds for sufficiently large τ. We also note that since $T \in (0,1)$, $0.88 < \Gamma(1+T) < 1$. At large distances, $\tilde{\chi}_{ij} \gg \Gamma(1+T)$, we can use the approximation $e^{-x} \approx 1 - x$ in (8), to write

$$P(\tilde{\chi}_{ij}) \approx \frac{C}{\tilde{\chi}_{ij}^{1/T}} \propto \frac{1}{\tilde{\chi}_{ij}^{1/T}} \approx p(\tilde{\chi}_{ij}), \tag{10}$$

where $p(x)$ is given by (1), while $C = \Gamma(1+T)^{1/T}$, $0.56 < C < 1$. At small distances, $\tilde{\chi}_{ij} \ll \Gamma(1+T)$, the exponential in (8) is much smaller than one, and we can write $P(\tilde{\chi}_{ij}) \approx 1 \approx p(\tilde{\chi}_{ij})$. In other words, at both small and large effective distances $\tilde{\chi}_{ij}$, the connection probability in the time-aggregated network resembles the Fermi-Dirac connection probability in the $\mathbb{S}^1$ model. Fig. 1 illustrates this effect in the time-aggregated networks of synthetic counterparts of real systems, whose snapshots can also have different average degrees $\bar{k}_t, t = 1, \ldots, \tau$ (see "Methods").

Given this equivalence, in Fig. 2 we apply Mercator, a recently developed embedding method based on the $\mathbb{S}^1$ model[23], to the time-aggregated network of the synthetic counterparts of the hospital and primary school. Mercator infers the nodes' coordinates $(\tilde{\kappa}, \theta)$ from the time-aggregated network (see "Methods"), and from $\tilde{\kappa}$ we estimate κ using (6). We also modified Mercator to use the connection probability in (7) instead of the connection probability in (1) (see Supplementary Information, Sect. VI). Fig. 2 shows that the two versions of Mercator perform similarly, inferring the nodes' latent coordinates remarkably well. Similar results hold for the synthetic counterparts of the rest of the real systems (Supplementary Information, Sect. II). In the rest of the paper, we use the original version of Mercator as its implementation is simpler and does not require knowledge of parameter τ.

Aggregation interval. As the aggregation interval τ increases, the time-aggregated network becomes denser, eventually turning into a fully connected network. This can be seen in (5), where irrespective of network size, at $\tau \to \infty$, $P(\chi_{ij}) \to 1, \forall i, j$. Further, at $\tau \to \infty$, $\alpha \to \infty$, and by (9) $\tilde{\chi}_{ij} \to 0, \forall i, j$. Clearly, no meaningful inference can be made in a fully connected network as all nodes "look the same". Thus for an accurate inference of the nodes' coordinates the interval τ has to be sufficiently small such that the corresponding time-aggregated network is not too dense. On the other hand, for intervals τ that are not sufficiently large there may not be enough data to allow accurate inference, as network snapshots are often very sparse in human proximity systems, consisting of only a fraction of nodes (Table 1). This effect is illustrated in Fig. 3, where we quantify the difference between real and inferred coordinates as a function of τ in a synthetic counterpart of the primary school. We see in Fig. 3 that there is a wide range of adequately large τ values, e.g., $500 < \tau < 10000$, where the accuracy of
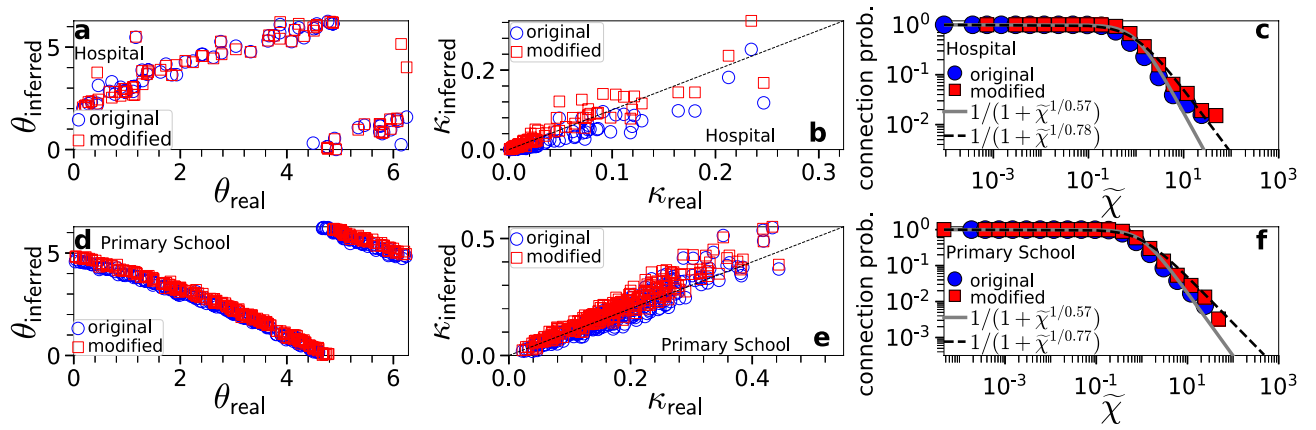
Figure 2. Inference of latent coordinates (κ, θ) with the original and modified versions of Mercator. The top row corresponds to a synthetic counterpart of the hospital, while the bottom row to a synthetic counterpart of the primary school. Both versions of Mercator are applied to the corresponding time-aggregated network formed over the full duration τ in Table 1. (**a,d**) Inferred versus real θ. (**b,e**) Inferred versus real κ. For each node, κ_{inferred} is estimated as $\kappa_{\text{inferred}} = \tilde{\kappa}/\alpha$, where $\tilde{\kappa}$ is the node's inferred latent degree in the time-aggregated network, while $\alpha = \tau^T / \Gamma(1 + T)$, with τ as in Table 1 and T as inferred by each version of Mercator. (**c,f**) Connection probability as a function of the effective distance $\tilde{\chi}$ in the time-aggregated network computed using the inferred coordinates $(\tilde{\kappa}, \theta)$. The solid grey and dashed black lines correspond to (1) with temperature T as inferred by each version of Mercator. For the two networks, the original version estimates $T = 0.57$, the modified version estimates $T = 0.78$ and 0.77, while the actual values are $T = 0.84$ and 0.72. In general, the modified version estimates values of T closer to the actual values. However, both versions of Mercator perform remarkably well at estimating the nodes' latent coordinates (κ, θ). We note that due to rotational symmetry of the model, the inferred angles can be globally shifted compared to the real angles by any value in $[0, 2\pi]$.
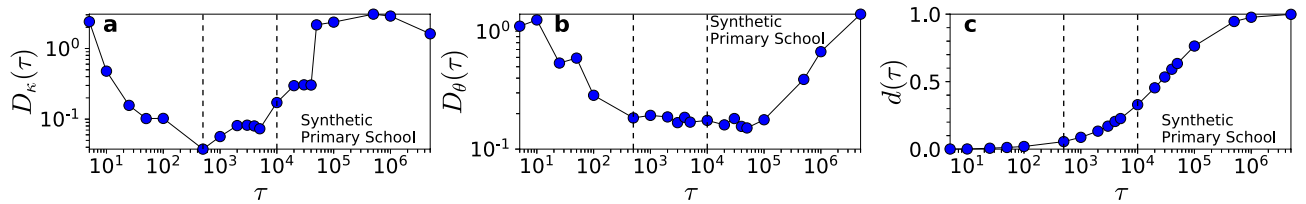

Figure 3. Inference accuracy vs. aggregation interval. The results correspond to a synthetic counterpart of the primary school constructed using the dynamic-$\mathbb{S}^1$ model. (**a**) Average difference between the inferred and real latent degrees as a function of the aggregation interval τ, $D_\kappa(\tau) = \sum_{i=1}^N |\kappa_{\text{inferred}}^i - \kappa_{\text{real}}^i|/N$, where $\kappa_{\text{inferred}}^i$ (κ_{real}^i) is the inferred (real) latent degree of node i. (**b**) Same as in (a) but for the average difference between the inferred and real angular coordinates, $D_\theta(\tau) = \sum_{i=1}^N |\theta_{\text{inferred}}^i - \theta_{\text{real}}^i|/N$. Before computing $D_\theta(\tau)$, the inferred angles are globally shifted such that the sum of the squared distances between real and inferred angles is minimized (to this end, we apply a Procrustean rotation[40], see Supplementary Information, Sect. VII for details). (**c**) Density of the time-aggregated network as a function of τ, $d(\tau) = 2L/[N(N-1)]$, where L is the number of links in the network. The vertical dashed lines indicate the interval $500 \leq \tau \leq 10000$. In this interval, $D_\kappa(\tau) < 0.2$, $D_\theta(\tau) < 0.2$, and $0.06 < d(\tau) < 0.33$.

inference for both κ and θ is simultaneously high, while as τ becomes too large or too small accuracy deteriorates. Similar results hold for the counterparts of the rest of the considered real systems (Supplementary Information, Sect. VII). The exact range of τ values where inference accuracy is high depends on the system's parameters, e.g., sparser networks (lower average snapshot degree) allow aggregation over longer intervals, as it takes longer for the time-aggregated network to become too dense. Further, our results with the synthetic counterparts suggest that daily aggregation intervals should be sufficient for accurate inference in most cases. Indeed, in this work we embed the time-aggregated networks of the considered real systems formed over the full observation durations τ in Table 1, as well as corresponding time-aggregated networks formed over individual observation days, obtaining in both cases meaningful results.

Hyperbolic maps of real systems. In Fig. 4 we apply Mercator to the time-aggregated network of the real networks in Table 1 and visualize the obtained hyperbolic maps and the corresponding connection probabilities. We see that the embeddings are meaningful, as we can identify in them actual node communities that correspond to groups of nodes located close to each other in the angular similarity space. These communities reflect the organization of students and teachers into classes (Fig. 4b,c), employees into departments (Fig. 4d), while no communities can be identified in the hospital (Fig. 4a). In all cases, we see a good match between empirical and theoretical connection probabilities (Fig. 4e–h). Next, we turn our attention to greedy routing.
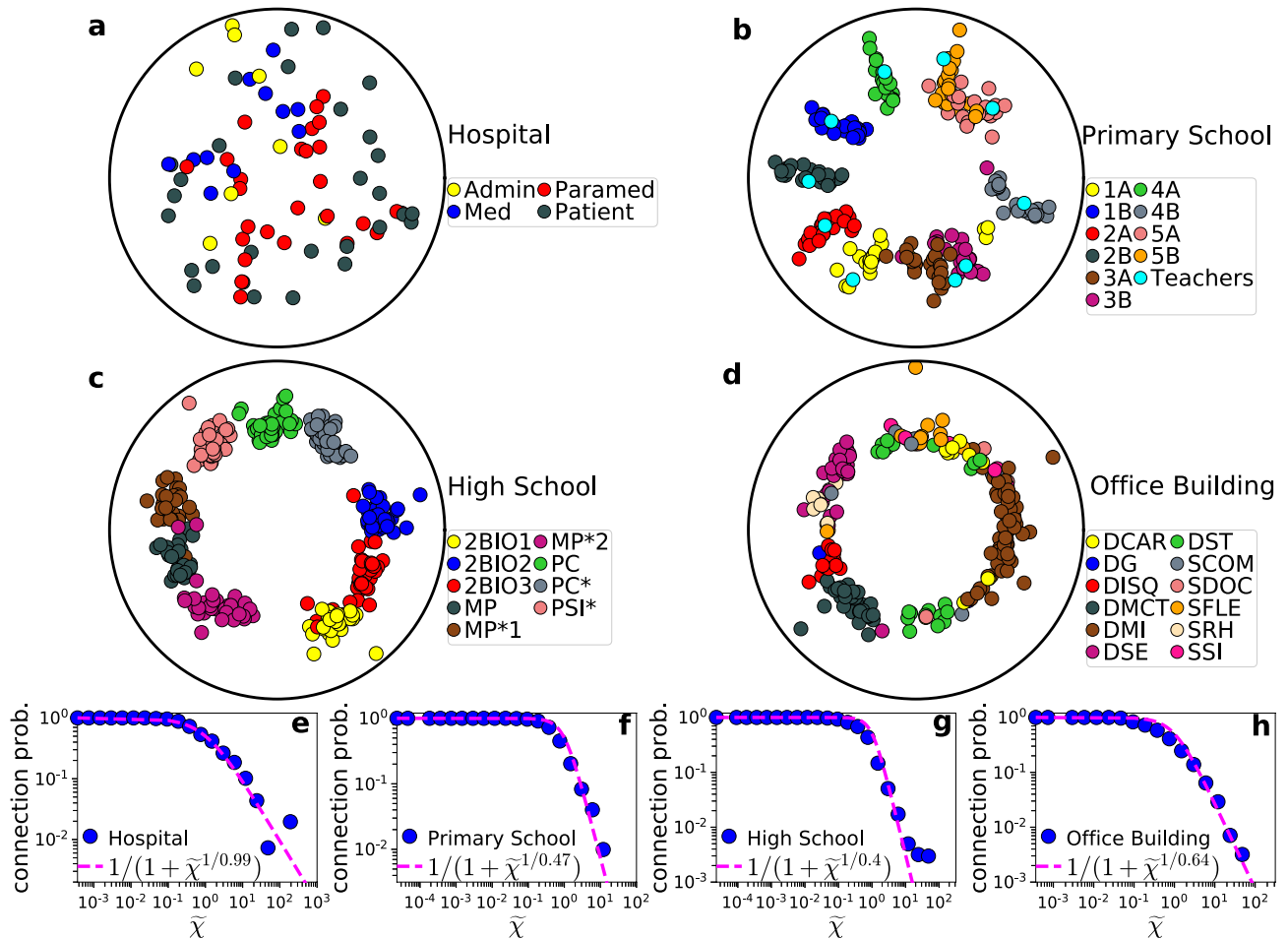
Figure 4. Hyperbolic embeddings of human proximity networks. (**a**-**d**) Hyperbolic maps of the time-aggregated networks of the hospital, primary school, high school and office building. In each case we consider the time-aggregated network formed over the full observation duration τ shown in Table 1. The nodes are positioned according to their inferred hyperbolic coordinates (r, θ) in the time-aggregated network [the radial coordinates r are computed using (4)]. The nodes are colored according to group membership information available in the metadata of each network. In the hospital, the nodes are administrative staff (Admin), medical doctors (Med), nurses and nurses' aides (Paramed), and patients (Patient). In the primary school, the nodes are teachers and students of the following classes: 1st grade (1A, 1B), 2nd grade (2A, 2B), 3rd grade (3A, 3B), 4th grade (4A, 4B), and 5th grade (5A, 5B). In the high school, the nodes are students of nine different classes with the following specializations: biology (2BIO1, 2BIO2, 2BIO3), mathematics and physics (MP, MP*1, MP*2), physics and chemistry (PC, PC*), and engineering studies (PSI*). In the office building, the nodes are employees working in different departments such as scientific direction (DISQ), chronic diseases and traumatisms (DMCT), department of health and environment (DSE), human resources (SRH), and logistics (SFLE). (**e**-**h**) Corresponding empirical connection probabilities as a function of the effective distance $\tilde{\chi}$. The pink dashed lines correspond to (1) with temperatures T as inferred by Mercator, $T = 0.99, 0.47, 0.40$ and 0.64, respectively. The maps for the conference and Friends & Family can be found in Supplementary Information, Sect. III. Daily hyperbolic maps for each real system can be found in Supplementary Information, Sect. V.

Human-to-human greedy routing. A problem of significant interest in mobile networking is how to efficiently route data in opportunistic networks, like human proximity systems, where the mobility of nodes creates contact opportunities among nodes that can be used to connect parts of the network that are otherwise disconnected[1–3,41]. Motivated by this problem, and by the remarkable efficiency of hyperbolic greedy routing in traditional complex networks[26,33,35], we investigate here if hyperbolic greedy routing can facilitate navigation in human proximity systems. To this end, we consider the following simplest greedy routing process, which performs routing on the temporal network using the coordinates inferred from the time-aggregated network.

Human-to-human greedy routing (H2H-GR). In H2H-GR, a node's address is its coordinates $(\tilde{\kappa}, \theta)$, and each node knows its own address, the addresses of its neighbors (nodes currently within proximity range), and the destination address written in the packet. A node holding the packet (carrier) forwards the packet to its neighbor with the smallest effective distance to the destination, but only if that distance is smaller than the distance between the carrier and the destination. Otherwise, or if the carrier currently has no neighbors, the carrier

6

Real network	H2H-GR (current coordinates)	H2H-GR (previous coordinates)	H2H-RR
Hospital	$p_s = 0.80, \bar{s} = 2.2$	$p_s = 0.47, \bar{s} = 2.0$	$p_s = 0.38, \bar{s} = 7.0$
Primary school	$p_s = 0.82, \bar{s} = 3.9$	$p_s = 0.65, \bar{s} = 3.6$	$p_s = 0.34, \bar{s} = 24.9$
Conference	$p_s = 0.70, \bar{s} = 2.2$	$p_s = 0.35, \bar{s} = 2.0$	$p_s = 0.29, \bar{s} = 7.9$
High school	$p_s = 0.29, \bar{s} = 2.0$	$p_s = 0.13, \bar{s} = 1.9$	$p_s = 0.07, \bar{s} = 5.9$
Office building	$p_s = 0.15, \bar{s} = 1.4$	$p_s = 0.10, \bar{s} = 1.4$	$p_s = 0.06, \bar{s} = 2.5$
Friends and family	$p_s = 0.45, \bar{s} = 1.8$	$p_s = 0.31, \bar{s} = 2.0$	$p_s = 0.21, \bar{s} = 5.3$

Table 2. Success ratio p_s and average stretch $\bar{s}$ of H2H-GR and H2H-RR in real networks. H2H-GR uses the coordinates inferred either from the time-aggregated network of the considered day where routing is performed (current coordinates); or from the time-aggregated network of the previous day (previous coordinates). The considered days in the hospital, primary school, conference, high school and office building are observation days 5, 2, 3, 5 and 10, respectively. In Friends & Family, the considered day is the 31st of March 2011. For a fair comparison with H2H-GR that uses the previous coordinates, we ignore during all routing processes the nodes that exist in the considered day but not in the previous day, since for such nodes we cannot infer their coordinates from the previous day. The percentage of such nodes is 17%, 3%, 7%, 6%, 14% and 3% for the hospital, primary school, conference, high school, office building and Friends & Family, respectively. In all cases, routing is performed among all possible source-destination pairs in the considered day that also exist in the previous day.

keeps the packet. Clearly, a carrier delivers the packet to the destination if the latter is its neighbor. We note that there are no routing loops in H2H-GR, i.e., no node receives the same packet twice. Indeed, consider for instance a packet from a node i_0 to a node i_n, which has followed the path $\{i_0, i_1, i_2, \ldots, i_{n-1}, i_n\}$. This means that $\tilde{\chi}_{i_0 i_n} > \tilde{\chi}_{i_1 i_n} > \tilde{\chi}_{i_2 i_n} > \ldots > \tilde{\chi}_{i_{n-1} i_n}$, where $\tilde{\chi}_{i_k i_n}$ is the effective distance between nodes i_k and i_n. A node i_k in the path never forwards the packet to a node i_l with $l < k$, i.e., to a node that has seen the packet before, because $\tilde{\chi}_{i_j i_n} > \tilde{\chi}_{i_k i_n}$. We also note that in the thermodynamic limit ($N \to \infty$), there is a non-zero probability that a packet constantly moves closer to the destination but never actually reaches it. This event could theoretically occur at $N \to \infty$, as there could be a countably infinite number of intermediate nodes with asymptotically closer effective distances to the destination. In reality such event can never occur since the number of nodes N is bounded.

For each network in Table 1, we simulate H2H-GR in one of its observation days. We consider the following two cases: i) H2H-GR that uses the nodes' coordinates inferred from the time-aggregated network of the considered day (current coordinates); and ii) H2H-GR that uses the nodes' coordinates inferred from the time-aggregated network of the previous day (previous coordinates). In the time-aggregated network of a day, two nodes are connected if they are connected in at least one network snapshot in the day. We compare these two cases to a baseline *random routing strategy (H2H-RR)*, where the carrier first determines the set of its neighbors that have never received the packet before, and then forwards the packet to one of these neighbors at random. If the destination is a neighbor the carrier forwards the packet to it. The carrier keeps the packet if it currently has no neighbors, or if all of its neighbors have received the packet before. Thus, there are no routing loops in H2H-RR either.

Performance metrics. We evaluate the performance of the algorithms according to the following two metrics: (i) the percentage of successful paths, p_s, which is the proportion of paths that reach their destinations by the end of the considered day; and (ii) the average stretch over the successful paths, $\bar{s}$. We define the stretch as the ratio of the hop-lengths of the paths found by the algorithms to the corresponding shortest time-respecting paths[42] in the network.

The results are shown in Table 2. We see that H2H-GR that uses the current coordinates significantly outperforms H2H-RR in both success ratio and stretch. The improvement can be quite significant. For instance, in the primary school the success ratio increases from 34% to 82%, while the average stretch decreases from 24.9 to 3.9. Similarly, in the hospital the success ratio increases from 38% to 80%, while the average stretch decreases from 7 to 2.2. These results show that hyperbolic greedy routing can significantly improve navigation. However, the success ratio decreases considerably if H2H-GR uses the previous coordinates. This suggests that the node coordinates change to a considerable extend from one day to the next. In Supplementary Information, Sect. V, we verify that this is indeed the case. Nevertheless, H2H-GR that uses the previous coordinates still outperforms H2H-RR with respect to success ratio, while achieving significantly lower stretch similar to the stretch with the current coordinates (Table 2).

Table 3 shows the same results for the synthetic counterparts of the real systems, where we can make qualitatively similar observations. Further, we see that H2H-GR achieves higher success ratios using the inferred coordinates in the counterparts compared to the real systems. This is not surprising as the counterparts are by construction maximally congruent with the assumed geometric model (dynamic-$\mathbb{S}^1$). Also, H2H-GR that uses the previous coordinates maintains high success ratios in the counterparts. This is expected, as the coordinates in the counterparts do not change over time. Thus the coordinates inferred from the time-aggregated network of the previous day are quite similar (but not exactly the same) to the ones inferred from the time-aggregated network of the day where routing is performed (see Supplementary Information, Sect. V).

Synthetic network	H2H-GR (current coordinates)	H2H-GR (previous coordinates)	H2H-RR
Hospital	$p_s = 0.92, \bar{s} = 2.2$	$p_s = 0.78, \bar{s} = 2.2$	$p_s = 0.42, \bar{s} = 9.2$
Primary school	$p_s = 0.98, \bar{s} = 3.7$	$p_s = 0.97, \bar{s} = 3.8$	$p_s = 0.53, \bar{s} = 33.9$
Conference	$p_s = 0.85, \bar{s} = 2.4$	$p_s = 0.70, \bar{s} = 2.4$	$p_s = 0.31, \bar{s} = 9.8$
High school	$p_s = 0.72, \bar{s} = 2.7$	$p_s = 0.59, \bar{s} = 2.4$	$p_s = 0.11, \bar{s} = 7.8$
Office building	$p_s = 0.26, \bar{s} = 1.5$	$p_s = 0.17, \bar{s} = 1.5$	$p_s = 0.06, \bar{s} = 3.0$
Friends and family	$p_s = 0.82, \bar{s} = 2.2$	$p_s = 0.70, \bar{s} = 2.3$	$p_s = 0.23, \bar{s} = 5.4$

Table 3. Same as in Table 2 but for the synthetic counterparts of the real systems constructed with the dynamic-$\mathbb{S}^1$ model. The results in each case correspond to one temporal network realization, while H2H-GR uses inferred coordinates as in Table 2.
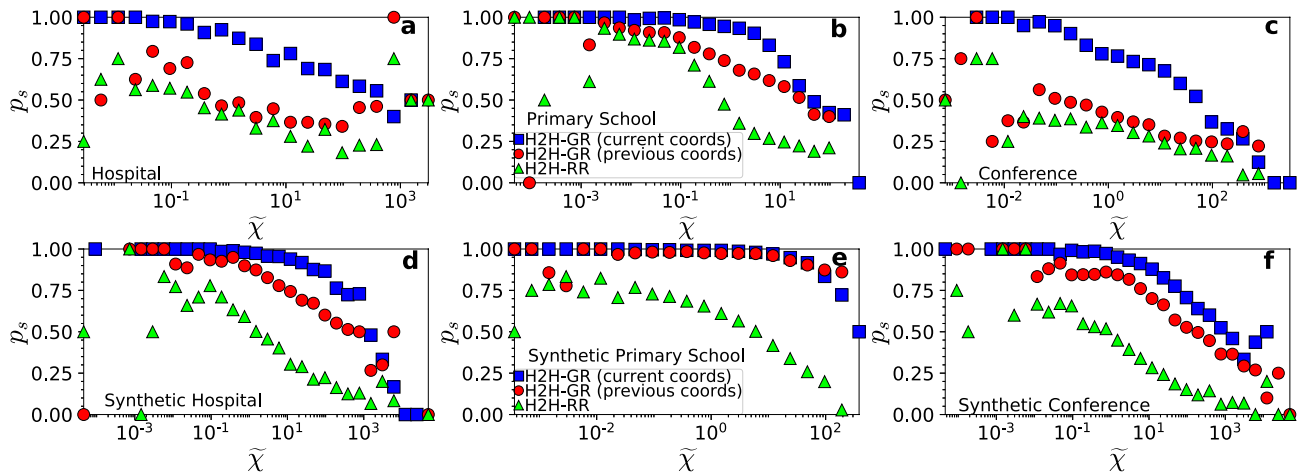
Figure 5. Success ratio p_s of H2H-GR and H2H-RR as a function of the effective distance $\tilde{\chi}$ between source-destination pairs. The top row corresponds to the results of the hospital, primary school and conference in Table 2, while the bottom row to the results of their synthetic counterparts in Table 3. The success ratio for H2H-RR and H2H-GR that uses the previous coordinates is shown as a function of the effective distance between the pairs in the previous day. Similar results hold for the other real networks and their synthetic counterparts (Supplementary Information, Sect. IV).
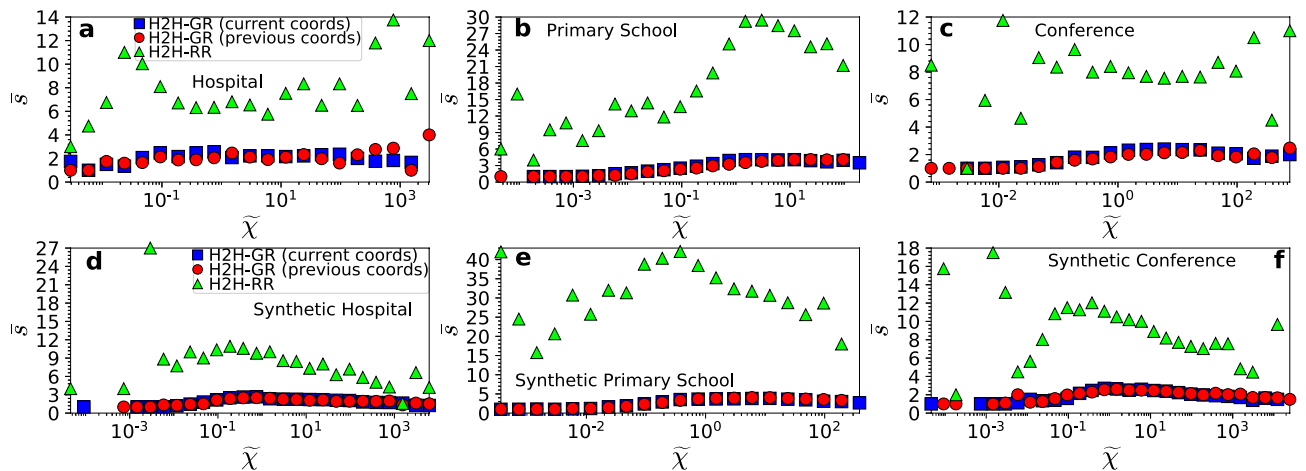

Figure 6. Same as in Fig. 5 but for the average stretch $\bar{s}$. Similar results hold for the other real networks and their synthetic counterparts (Supplementary Information, Sect. IV).

The metrics in Tables 2 and 3 are computed across all source-destination pairs. In Figs. 5 and 6 we also compute these metrics as a function of the effective distance between the source-destination pairs. We see that H2H-GR that uses the current coordinates achieves high success ratios, approaching 100%, as the effective distance between the pairs decreases. As the effective distance between the pairs increases, the success ratio decreases. The average stretch for successful H2H-GR paths is always low.

Network	AUROC real	AUPR real	AUROC chance	AUPR chance	AUROC synthetic	AUPR synthetic
Hospital	0.78	0.70	0.5	0.43	0.90	0.77
Primary school	0.81	0.62	0.5	0.20	0.87	0.71
Conference	0.66	0.34	0.5	0.22	0.88	0.62
High school	0.89	0.40	0.5	0.05	0.94	0.59
Office building	0.71	0.12	0.5	0.05	0.90	0.41
Friends and family	0.86	0.60	0.5	0.10	0.93	0.72

Table 4. AUROC and AUPR for geometric link prediction in real networks and their synthetic counterparts. The day of interest is day 3 in the hospital and day 2 in the rest of the networks. Geometric link prediction uses the effective distances among the nodes inferred from the time-aggregated network of the previous day. "AUPR chance" corresponds to link prediction based on pure chance in the real networks. It equals the ratio of the number of connected pairs to the total number of pairs in the time-aggregated network of the day of interest. AUPR chance values for the synthetic counterparts are similar as in the real networks and not shown for brevity.

H2H-RR also achieves considerably high success ratios for pairs separated by small distances (Fig. 5). This is because, even though packets in H2H-RR are forwarded to neighbors at random, the neighbors are not random nodes but nodes closer to the carriers in the hyperbolic space. Thus, packets between pairs separated by smaller distances have higher chances of finding their destinations. However, the stretch of successful paths in H2H-RR is quite high (Fig. 6). Further, we see that in real networks the success ratio of H2H-GR that uses the previous coordinates resembles in most cases the one of H2H-RR (Fig. 5a–c and Supplementary Fig. S4). However, the stretch in H2H-GR is always significantly lower than in H2H-RR (Figs. 6a–c and Supplementary Fig. S5).

Taken altogether, these results show that hyperbolic greedy routing can facilitate efficient navigation in human proximity networks. The success ratio for pairs separated by large effective distances can be low (Fig. 5). However, it is possible that more sophisticated algorithms than the one considered here could improve the success ratio for such pairs without significantly sacrificing stretch. Further, using coordinates from past embeddings decreases the success ratio. Even though the average stretch remains low, this observation suggests that the evolution of the nodes' coordinates should also be taken into account. Such investigations are beyond the scope of this paper. Finally, we note that in Supplementary Information, Sect. IV, we consider H2H-GR that uses only the angular similarity distances among the nodes, and find that it performs worse than H2H-GR that uses the effective distances. This means that in addition to node similarities, node expected degrees (or popularities[31]) also matter in H2H-GR, even though the distribution of node degrees in human proximity systems is quite homogeneous[22].

Link prediction. In this section, we turn our attention to link prediction. We want to see how well we can predict if two nodes are connected in the time-aggregated network of a day, if we know the effective distances among the nodes in the previous day. To this end, for each pair of nodes i, j in the previous day that is also present in the day of interest, we assign a score $s_{ij} = 1/\tilde{\chi}_{ij}$, where $\tilde{\chi}_{ij}$ is the inferred effective distance between i and j in the time-aggregated network of the previous day. The higher the s_{ij}, the higher is the likelihood that i and j are connected in the day of interest. We call this approach *geometric*. To quantify the quality of link prediction, we use two standard metrics: (i) the *Area Under the Receiver Operating Characteristic curve (AUROC)*; and (ii) the *Area Under the Precision-Recall curve (AUPR)*[43]. These metrics are described below.

The AUROC represents the probability that a randomly selected connected pair of nodes is given a higher score than a randomly selected disconnected pair of nodes in the day of interest. The degree to which the AUROC exceeds 0.5 indicates how much better the method performs than pure chance. As the name suggests, the AUROC is equal to the total area under the *Receiver Operating Characteristic (ROC)* curve. To compute the ROC curve, we order the pairs of nodes in the descending order of their scores, from the largest s_{ij} to the smallest s_{ij}, and consider each score to be a threshold. Then, for each threshold we calculate the fraction of connected pairs that are above the threshold (i.e., the True Positive Rate TPR) and the fraction of disconnected pairs that are above the threshold (i.e., the False Positive Rate FPR). Each point on the ROC curve gives the TPR and FPR for the corresponding threshold. When representing the TPR in front of the FPR, a totally random guess would result in a straight line along the diagonal $y = x$, while the degree by which the ROC curve lies above the diagonal indicates how much better the algorithm performs than pure chance. AUROC = 1 means a perfect classification (ordering) of the pairs, where the connected pairs are placed in the top of the ordered list.

The AUPR represents how accurately the method can classify pairs of nodes as connected and disconnected based on their scores. It is equal to the total area under the *Precision-Recall (PR)* curve. To compute the PR curve, we again order the pairs of nodes in the descending order of their scores, and consider each score to be a threshold. Then, for each threshold we calculate the TPR, which is called Recall, and the Precision, which is the fraction of pairs above the threshold that are connected. Each point on the PR curve gives the Precision and Recall for the corresponding threshold. A random guess corresponds to a straight line parallel to the Recall axis at the level where Precision equals the ratio of the number of connected pairs to the total number of pairs. The higher the AUPR the better the method is, while a perfect classifier yields AUPR = 1.

The results for the considered real networks and their synthetic counterparts are shown in Table 4. The corresponding ROC and PR curves are shown in Fig. 7. We see that geometric link prediction significantly outperforms
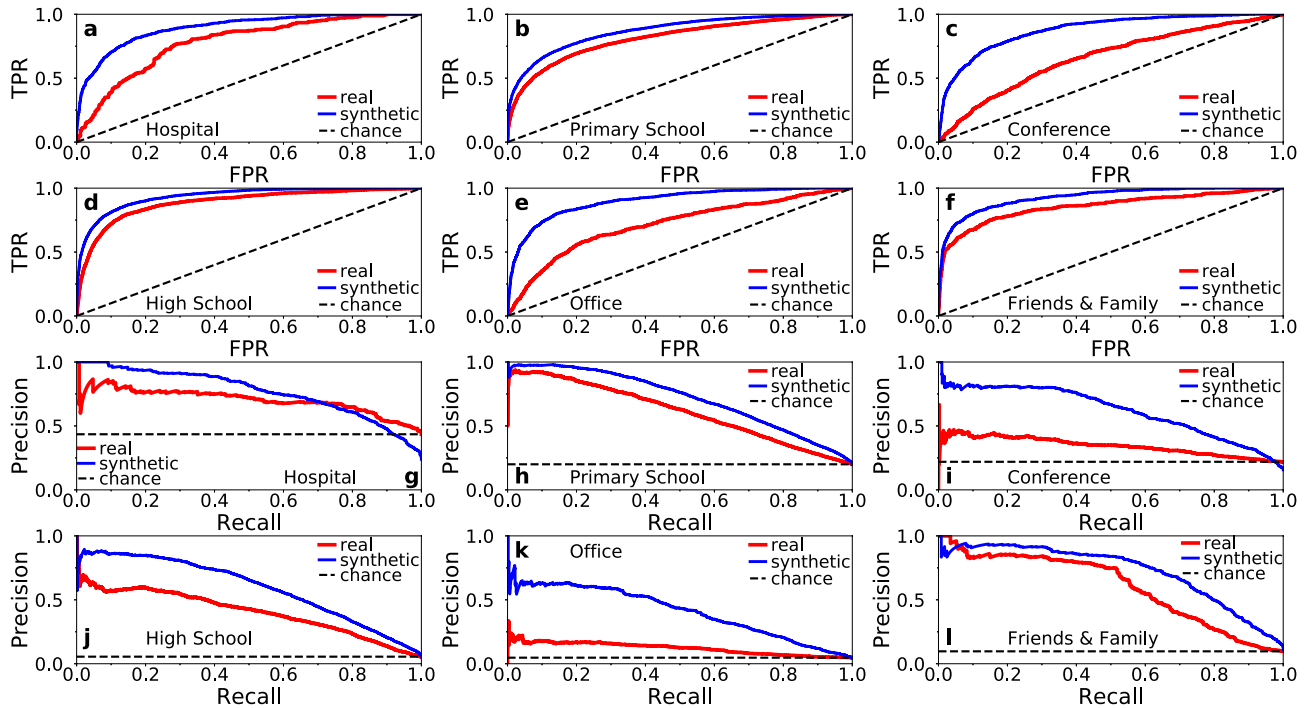
Figure 7. ROC and PR curves for geometric link prediction in real networks and their synthetic counterparts. (**a**–**f**) show the ROC curves, while (**g**–**l**) the PR curves, corresponding to the results in Table 4. The dashed black lines correspond to link prediction based on chance; these lines in (**g**–**l**) correspond to the AUPR chance values in Table 4.

Network	AUROC real	AUPR real	AUROC synthetic	AUPR synthetic
Hospital	0.75	0.79	0.85	0.69
Primary school	0.79	0.52	0.84	0.62
Conference	0.67	0.37	0.85	0.57
High school	0.88	0.44	0.89	0.52
Office building	0.73	0.10	0.86	0.35
Friends and family	0.85	0.54	0.89	0.64

Table 5. Same as in Table 4 but for the CN approach.

chance in all cases. These results constitute another validation that the embeddings are meaningful, and illustrate that they have significant predictive power. As can be seen in Table 4 and Fig. 7, link prediction is more accurate in the synthetic counterparts. This is again expected since the counterparts are by construction maximally congruent with the underlying geometric space, while the node coordinates in them do not change over time.

We also compute the same metrics as in Table 4 but for a simple heuristic, where the score s_{ij} between two nodes i and j is the number of common neighbors they have in the time-aggregated network of the previous day (CN approach). The results are shown in Table 5. Interestingly, we see that the performance of the geometric and CN approaches is quite similar in real networks, suggesting that the latter is a good heuristic for link prediction in human proximity systems. The performance of the two approaches is also positively correlated in the synthetic counterparts (Tables 4 and 5). This is expected since the smaller the effective distance between two nodes the larger is the *expected* number of common neighbors the nodes have. However, as can be seen in Tables 4 and 5, in the counterparts the geometric approach performs better than the CN approach. This suggests that the performance of the former could be further improved in real systems, if more accurate predictions of the node coordinates in the period of interest could be made.

Epidemic spreading.   Finally, we consider epidemic spreading. Here, predicting the arrival time of an epidemic is crucial for developing better containment measures for infectious diseases[44,45]. In the context of the global air transportation network, Brockmann and Helbing showed that the epidemic arrival time in a country can be well predicted by the effective distance between the country and the infection source country[45]. The effective distance between two countries is defined as the length of the shortest weighted path connecting the two
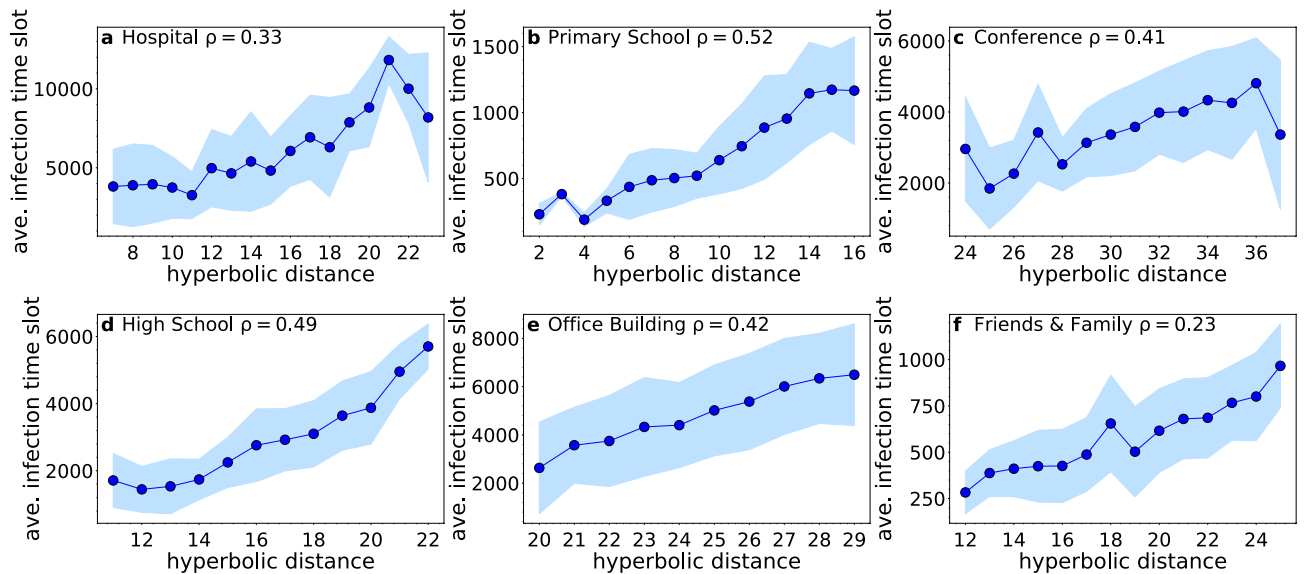
Figure 8. Average infection time slot as a function of the hyperbolic distance from the infected source node in real networks. In each case we consider the inferred hyperbolic distances in the time-aggregated network formed over the full observation duration. The hyperbolic distance is binned into bins of size $\delta = 1$ and the plots show the average infection time slot for nodes whose hyperbolic distance from the source node falls within each bin. The shaded area identifies the region corresponding to one standard deviation away from the average. Bins with less than 5 samples are ignored. The results are averaged over 10 simulated SI processes. Each process starts with a different infected source node selected at random, while the infection probability per time slot is $\beta = 0.05$. Each plot indicates the average Spearman rank correlation coefficient ρ between the infection time slot and the hyperbolic distance across the 10 SI processes. In these plots we consider the hyperbolic distance instead of the equivalent effective distance $\tilde{\chi}$, as the former is more convenient for binning purposes.

countries in the air transportation network, where the weight of a link is a decreasing function of the air traffic between the endpoints of the link[45].

In a similar vein, here we show that in human proximity networks, the epidemic arrival time, i.e., the time slot at which a node becomes infected, is positively correlated with the hyperbolic distance between the node and the infected source node in the time-aggregated network. [We note that while in Ref.[45] the effective distances are directly defined by observable (weighted) path lengths, the effective distances in our case are defined by the nodes' latent coordinates that manifest themselves indirectly via the nodes' connections and disconnections in the (unweighted) time-aggregated network.] To this end, we consider the Susceptible-Infected (SI) epidemic spreading model[46]. In the SI, each node can be in one of two states, susceptible (S) or infected (I). At any time slot infected nodes infect susceptible nodes with whom they are within proximity range, with probability β. Thus, the transition of states is S→I. To simulate the SI process on temporal networks we use the dynamic SI implementation of the Network Diffusion library[47].

Figures 8 and 9 show the results for the considered real networks and their synthetic counterparts, respectively. We see that the epidemic arrival times are significantly correlated with the hyperbolic distance from the infected source node. The correlation in each case is measured in terms of Spearman's rank correlation coefficient ρ (see "Methods"). These results indicate that hyperbolic embedding could provide a new perspective for understanding and predicting the behavior of epidemic spreading in human proximity systems. We leave further explorations for future work.

Conclusion

Individual snapshots of human proximity networks are often very sparse, consisting of a small number of interacting nodes. Nevertheless, we have shown that meaningful hyperbolic embeddings of such systems are still possible. Our approach is based on embedding the time-aggregated network of such systems over an adequately large observation period, using mapping methods developed for traditional complex networks. We have justified this approach by showing that the connection probability in the time-aggregated network is compatible with the Fermi-Dirac connection probability in random hyperbolic graphs, on which existing embedding methods are based. From an applications' perspective, we have shown that the hyperbolic maps of real proximity systems can be used to identify communities, facilitate efficient greedy routing on the temporal network, and predict future links. Further, we have shown that epidemic arrival times in the temporal network are positively correlated with the distance from the infection sources in the maps. Overall, our work opens the door for a geometric description of human proximity systems.

Our results indicate that the node coordinates change over time in the hyperbolic spaces of human proximity networks. An interesting yet challenging future work direction is to identify the stochastic differential equations that dictate this motion of nodes. Such equations would allow us to make predictions about the future positions of nodes in their hyperbolic spaces over different timescales. This, in turn, could allow us to improve the
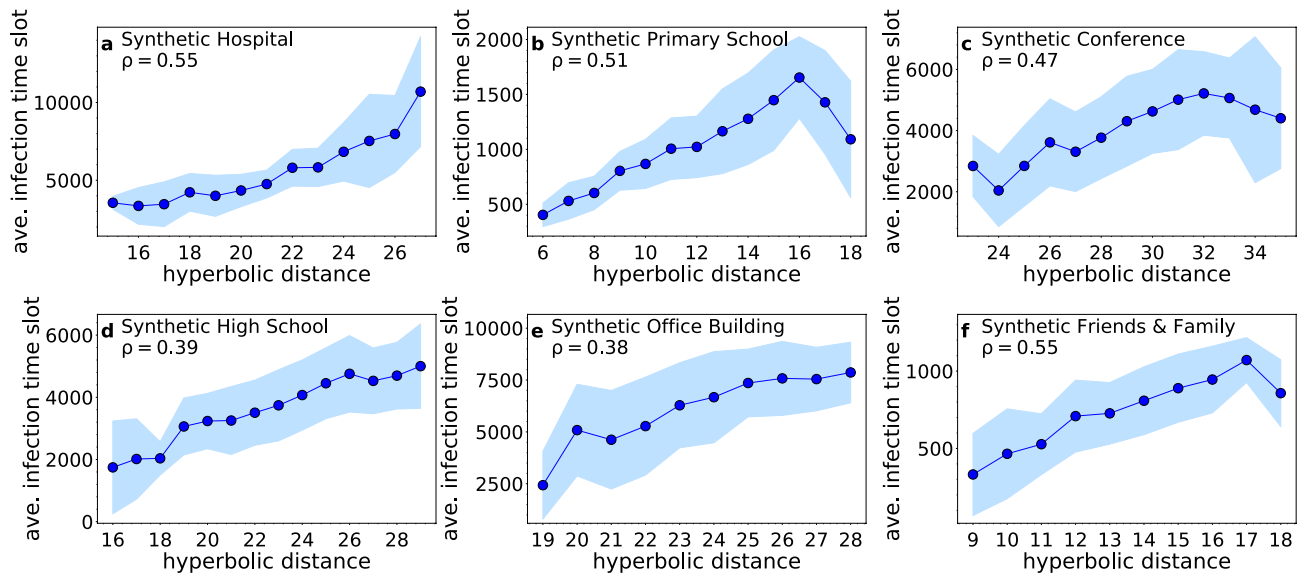
Figure 9. Same as in Fig. 8 but for the synthetic counterparts (using inferred hyperbolic distances).
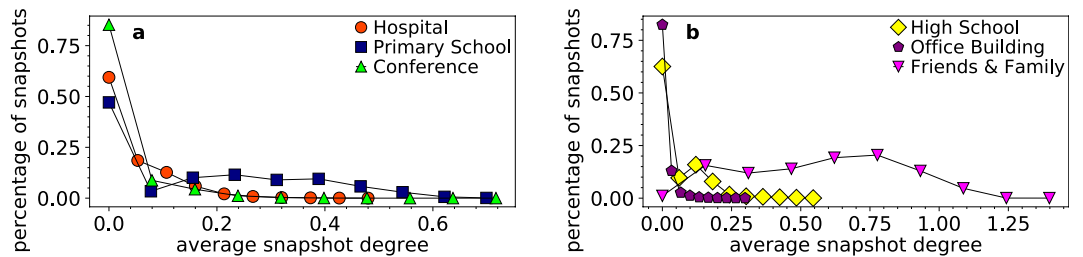

Figure 10. Distribution of the average snapshot degree in the considered real networks.

performance of tasks such as greedy routing and link prediction. This problem is relevant not only for human proximity systems, but for all complex networks where the hyperbolic node coordinates are expected to change over time, such as in social networks and the Internet[28]. Another problem is to extend existing hyperbolic embedding methods so that they can refine the nodes' coordinates on a snapshot-by-snapshot basis as new snapshots become available, without having to recompute each time a new embedding from scratch. Such methods could be based on the idea that a local change in the system (new connections or disconnections) should involve mostly the neighborhood (coordinates of the nodes) around the change. For this purpose, techniques based on quadtree structures as in Ref.[48] appear promising. Further, one might want to penalize large displacements based on the idea that the coordinates should be changing gradually from snapshot to snapshot. To this end, Gaussian transition models for the coordinates as in Ref.[37] seem appropriate. Methods for dynamic embedding in hyperbolic spaces should be useful not only for human proximity systems, but for temporal networks in general.

Methods

Generating synthetic networks with the dynamic-$\mathbb{S}^1$ model.

For each real network we construct its synthetic counterpart using the dynamic-$\mathbb{S}^1$ model as in Ref.[22]. Specifically, each counterpart has the same number of nodes N and total duration (number of time slots) τ as the corresponding real network in Table 1, while the latent variable κ_i of each node $i = 1, \ldots, N$ is set equal to the node's average degree per slot in the real network. The average degree $\bar{k}_t$ in each snapshot G_t, $t = 1, \ldots, \tau$, is set equal to the average degree in the corresponding real snapshot at slot t—Fig. 10 shows the distribution of $\bar{k}_t$. Finally, the temperature T is set such that the resulting average time-aggregated degree, $\bar{k}$, is similar to the one in the real network. Each "day" in each counterpart corresponds to the same time slots as the corresponding day in the real system. See Ref.[22] for further details.

Mercator.

Mercator[23] combines the Laplacian Eigenmaps (LE) approach of Ref.[36] with maximum likelihood estimation (MLE) to produce fast and accurate embeddings. It can embed networks with arbitrary degree distributions. In a nutshell, Mercator takes as input the network's adjacency matrix. It infers the nodes' latent degrees ($\tilde{\kappa}$) using the nodes' observed degrees in the network and the connection probability in the $\mathbb{S}^1$ model. To infer the nodes' angular coordinates (θ), Mercator first utilizes the LE approach adjusted to the $\mathbb{S}^1$ model, in order to determine initial angular coordinates for the nodes. These initial angular coordinates are then refined using MLE, which adjusts the angular coordinates by maximizing the probability that the given network is produced

by the $\mathbb{S}^1$ model. Mercator also estimates the value of the temperature parameter T. The code implementing Mercator is made publicly available by the authors of[23] at https://github.com/networkgeometry/mercator. We have used the code as is without any modifications.

As mentioned in the main text, we also considered a modified version of Mercator that replaces the connection probability of the $\mathbb{S}^1$ model in (1) with the connection probability in (7). This modification requires several changes to the original Mercator implementation that we describe in Supplementary Information, Sect. VI.

Epidemic arrival time and hyperbolic distance correlation. To quantify the correlation between the time slot at which a node becomes infected and its hyperbolic distance from the infected source node, we use Spearman's rank correlation coefficient ρ[49]. Formally, given n values X_i, Y_i, the values are converted to ranks rg_{X_i}, rg_{Y_i}, and Spearman's ρ is computed as

$$\rho = \frac{\text{cov}(rg_X, rg_Y)}{\sigma_{rg_X}\sigma_{rg_Y}}, \tag{11}$$

where $\text{cov}(rg_X, rg_Y)$ is the covariance of the rank variables, while $\sigma_{rg_X}, \sigma_{rg_Y}$ are the standard deviations of the rank variables. Spearman's ρ takes values between -1 and 1, and assesses monotonic relationships. $\rho = 1$ ($\rho = -1$) occurs when there is a perfect monotonic increasing (decreasing) relationship between variables X and Y, while $\rho = 0$ indicates that there is no tendency for Y to either increase or decrease when X increases.

References

1. Hui, P. *et al.* Pocket switched networks and human mobility in conference environments. In *Proceedings of the ACM SIGCOMM Workshop on Delay-tolerant Networking*, WDTN 05, 244–251, https://doi.org/10.1145/1080139.1080142 (ACM, New York, USA, 2005).
2. Chaintreau, A. *et al.* Impact of human mobility on opportunistic forwarding algorithms. *IEEE Trans. Mob. Comput.* **6**, 606–620. https://doi.org/10.1109/TMC.2007.1060 (2007).
3. Karagiannis, T., Le Boudec, J.-Y. & Vojnovic, M. Power law and exponential decay of intercontact times between mobile devices. *IEEE Trans. Mob. Comput.* **9**, 1377–1390. https://doi.org/10.1109/TMC.2010.99 (2010).
4. Dong, W., Lepri, B. & Pentland, A. Modeling the co-evolution of behaviors and social relationships using mobile phone data. In *Proceedings of the International Conference on Mobile and Ubiquitous Multimedia*, MUM '11, 134–143, https://doi.org/10.1145/2107596.2107613 (ACM, New York, USA, 2011).
5. Aharony, N., Pan, W., Ip, C., Khayal, I. & Pentland, A. Social fMRI: investigating and shaping social mechanisms in the real world. *Pervasive Mob. Comput.* **7**, 643–659. https://doi.org/10.1016/j.pmcj.2011.09.004 (2011).
6. Barrat, A. & Cattuto, C. Face-to-face interactions. In *Social Phenomena: From Data Analysis to Models*, 37–57, https://doi.org/10.1007/978-3-319-14011-7_3 (Springer, Cham, 2015).
7. Holme, P. Temporal network structures controlling disease spreading. *Phys. Rev. E* **94**, 022305. https://doi.org/10.1103/PhysRevE.94.022305 (2016).
8. Holme, P. & Litvak, N. Cost-efficient vaccination protocols for network epidemiology. *PLOS Comput. Biol.* **13**, 1–18. https://doi.org/10.1371/journal.pcbi.1005696 (2017).
9. Isella, L. *et al.* Whats in a crowd? Analysis of face-to-face behavioral networks. *J. Theor. Biol.* **271**, 166–180. https://doi.org/10.1016/j.jtbi.2010.11.033 (2011).
10. Stehlé, J. *et al.* High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS ONE* **6**, e23176. https://doi.org/10.1371/journal.pone.0023176 (2011).
11. Vanhems, P. *et al.* Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLoS ONE* **8**, e73970. https://doi.org/10.1371/journal.pone.0073970 (2013).
12. Mastrandrea, R., Fournet, J. & Barrat, A. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLoS ONE* **10**, e0136497. https://doi.org/10.1371/journal.pone.0136497 (2015).
13. Génois, M. *et al.* Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers. *Netw. Sci.* **3**, 326–347. https://doi.org/10.1017/nws.2015.10 (2015).
14. Sociopatterns. http://www.sociopatterns.org/.
15. Henderson, T., Kotz, D. & Abyzov, I. The changing usage of a mature campus-wide wireless network. *Comput. Netw.* **52**, 2690–2712. https://doi.org/10.1016/j.comnet.2008.05.003 (2008).
16. Starnini, M., Lepri, B., Baronchelli, A., Barrat, A., Cattuto, C. & Pastor-Satorras, R. Robust modeling of human contact networks across different scales and proximity-sensing techniques. In *Soc. Inform.* https://doi.org/10.1007/978-3-319-67217-5_32 (Springer, Cham, 2017).
17. Sekara, V., Stopczynski, A. & Lehmann, S. Fundamental structures of dynamic social networks. *Proc. Natl. Acad. Sci. USA* **113**, 9977–9982. https://doi.org/10.1073/pnas.1602803113 (2016).
18. Rodríguez-Flores, M. A. & Papadopoulos, F. Similarity forces and recurrent components in human face-to-face interaction networks. *Phys. Rev. Lett.* **121**, 258301. https://doi.org/10.1103/PhysRevLett.121.258301 (2018).
19. Starnini, M., Baronchelli, A. & Pastor-Satorras, R. Modeling human dynamics of face-to-face interaction networks. *Phys. Rev. Lett.* **110**, 168701. https://doi.org/10.1103/PhysRevLett.110.168701 (2013).
20. Serrano, M. A., Krioukov, D. & Boguñá, M. Self-similarity of complex networks and hidden metric spaces. *Phys. Rev. Lett.* **100**, 078701. https://doi.org/10.1103/PhysRevLett.100.078701 (2008).
21. Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A. & Boguñá, M. Hyperbolic geometry of complex networks. *Phys. Rev. E* **82**, 036106. https://doi.org/10.1103/PhysRevE.82.036106 (2010).
22. Papadopoulos, F. & Rodríguez-Flores, M. A. Latent geometry and dynamics of proximity networks. *Phys. Rev. E* **100**, 052313. https://doi.org/10.1103/PhysRevE.100.052313 (2019).
23. García-Pérez, G., Allard, A., Serrano, M. Á. & Boguñá, M. Mercator: uncovering faithful hyperbolic embeddings of complex networks. *New J. Phys.* **21**, 123033. https://doi.org/10.1088/1367-2630/ab57d2 (2019).
24. Génois, M. & Barrat, A. Can co-location be used as a proxy for face-to-face contacts?. *EPJ Data Sci.* **7**, 11. https://doi.org/10.1140/epjds/s13688-018-0140-1 (2018).

25. Boguñá, M. & Pastor-Satorras, R. Class of correlated random networks with hidden variables. *Phys. Rev. E* **68**, 036112. https://doi.org/10.1103/PhysRevE.68.036112 (2003).
26. Boguñá, M., Papadopoulos, F. & Krioukov, D. Sustaining the internet with hyperbolic mapping. *Nat. Commun.* **1**, 62 (2010).
27. Papadopoulos, F., Psomas, C. & Krioukov, D. Network mapping by replaying hyperbolic growth. *IEEE/ACM Trans. Netw.* **23**, 198–211. https://doi.org/10.1109/TNET.2013.2294052 (2015).
28. Papadopoulos, F., Aldecoa, R. & Krioukov, D. Network geometry inference using common neighbors. *Phys. Rev. E* **92**, 022807. https://doi.org/10.1103/PhysRevE.92.022807 (2015).
29. Alanis-Lobato, G., Mier, P. & Andrade-Navarro, M. A. Manifold learning and maximum likelihood estimation for hyperbolic network embedding. *Appl. Netw. Sci.* **1**, 10. https://doi.org/10.1007/s41109-016-0013-0 (2016).
30. Bläsius, T., Friedrich, T., Krohmer, A. & Laue, S. Efficient embedding of scale-free graphs in the hyperbolic plane. *IEEE/ACM Trans. Network.* **26**, 920–933. https://doi.org/10.1109/TNET.2018.2810186 (2018).
31. Papadopoulos, F., Kitsak, M., Serrano, M. Á., Boguñá, M. & Krioukov, D. Popularity versus similarity in growing networks. *Nature* **489**, 537–540 (2012).
32. Kleineberg, K.-K., Boguñá, M., Serrano, M. Á. & Papadopoulos, F. Hidden geometric correlations in real multiplex networks. *Nat. Phys.* **12**, 1076–1081. https://doi.org/10.1038/nphys3812 (2016).
33. Ortiz, E., Starnini, M. & Serrano, M. Á. Navigability of temporal networks in hyperbolic space. *Sci. Rep.* **7**, 15054. https://doi.org/10.1038/s41598-017-15041-0 (2017).
34. Alanis-Lobato, G., Mier, P. & Andrade-Navarro, M. A. The latent geometry of the human protein interaction network. *Bioinformatics* **34**, 2826–2834. https://doi.org/10.1093/bioinformatics/bty206 (2018).
35. Allard, A. & Serrano, M. Á. Navigable maps of structural brain networks across species. *PLOS Comput. Biol.* **16**, 1–20. https://doi.org/10.1371/journal.pcbi.1007584 (2020).
36. Muscoloni, A., Thomas, J. M., Ciucci, S., Bianconi, G. & Cannistraci, C. V. Machine learning meets complex networks via coalescent embedding in the hyperbolic space. *Nat. Commun.* **8**, 1615. https://doi.org/10.1038/s41467-017-01825-5 (2017).
37. Kim, B., Lee, K. H., Xue, L. & Niu, X. A review of dynamic network models with latent variables. *Stat. Surv.* **12**, 105–135. https://doi.org/10.1214/18-SS121 (2018).
38. Cui, P., Wang, X., Pei, J. & Zhu, W. A survey on network embedding. *IEEE Trans. Knowl. Data Eng.* **31**, 833–852. https://doi.org/10.1109/TKDE.2018.2849727 (2019).
39. Torricelli, M., Karsai, M. & Gauvin, L. weg2vec: Event embedding for temporal networks. *Sci. Rep.* **10**, 7164. https://doi.org/10.1038/s41598-020-63221-2 (2020).
40. Rohlf, F. J. & Slice, D. Extensions of the procrustes method for the optimal superimposition of landmarks. *Syst. Biol.* **39**, 40–59. https://doi.org/10.2307/2992207 (1990).
41. Conti, M. & Giordano, S. Mobile ad hoc networking: milestones, challenges, and new research directions. *IEEE Commun. Mag.* **52**, 85–96. https://doi.org/10.1109/MCOM.2014.6710069 (2014).
42. Holme, P. & Saramäki, J. (eds.) *Temporal Network Theory* (Springer International Publishing, 2019), 1 edn.
43. Saito, T. & Rehmsmeier, M. Precrec: fast and accurate precision-recall and ROC curve calculations in R. *Bioinformatics* **33**, 145–147. https://doi.org/10.1093/bioinformatics/btw570 (2016).
44. Gauvin, L., Panisson, A., Cattuto, C. & Barrat, A. Activity clocks: spreading dynamics on temporal networks of human contact. *Sci. Rep.* **3**, 3099 (2013).
45. Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342. https://doi.org/10.1126/science.1245200 (2013).
46. Kermack, W. O., McKendrick, A. G. & Walker, G. T. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A* **115**, 700–721. https://doi.org/10.1098/rspa.1927.0118 (1927).
47. Rossetti, G. *et al.* NDlib: a python library to model and analyze diffusion processes over complex networks. *Int. J. Data Sci. Anal.* **5**, 61–79. https://doi.org/10.1007/s41060-017-0086-6 (2018).
48. Looz, M. V. & Meyerhenke, H. Updating dynamic random hyperbolic graphs in sublinear time. *ACM J. Exp. Algorithmics* **23**, 1.6:1-1.6:30. https://doi.org/10.1145/3195635 (2018).
49. Spearman, C. The proof and measurement of association between two things. *Am. J. Psychol.* **15**, 72–101 (1904).

Acknowledgements

Author contributions

M.A.R.-F. and F.P. designed the research and wrote the manuscript; M.A.R.-F. conducted the research.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-77277-7.

Correspondence and requests for materials should be addressed to F.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.