The structure and information spread capability of the network formed by integrated fitness apps

Euodia Vermeulen
Department of Industrial Engineering, Faculty of Engineering, Stellenbosch University, Stellenbosch, South Africa, and
Sara Grobbelaar
Department of Industrial Engineering, Faculty of Engineering, Stellenbosch University, Stellenbosch, South Africa and DSI-NRF CoE in Scientometrics and Science, Technology and Innovation Policy, Stellenbosch University, Stellenbosch, South Africa

Received 10 January 2022
Revised 29 April 2022
Accepted 21 June 2022

Abstract

Purpose – In this article we aim to understand how the network formed by fitness tracking devices and associated apps as a subset of the broader health-related Internet of things is capable of spreading information.

Design/methodology/approach – The authors used a combination of a content analysis, network analysis, community detection and simulation. A sample of 922 health-related apps (including manufacturers' apps and developers) were collected through snowball sampling after an initial content analysis from a Google search for fitness tracking devices.

Findings – The network of fitness apps is disassortative with high-degree nodes connecting to low-degree nodes, follow a power-law degree distribution and present with low community structure. Information spreads faster through the network than an artificial small-world network and fastest when nodes with high degree centrality are the seeds.

Practical implications – This capability to spread information holds implications for both intended and unintended data sharing.

Originality/value – The analysis confirms and supports evidence of widespread mobility of data between fitness and health apps that were initially reported in earlier work and in addition provides evidence for the dynamic diffusion capability of the network based on its structure. The structure of the network enables the duality of the purpose of data sharing.

Keywords Information sharing, Data sharing, Network analysis, Ethics, E-Health, Human computer interaction

Paper type Research paper

1. Introduction

Healthcare is undergoing a digital transformation, enabled by digital technologies and the advances in data analytics, artificial intelligence, health-related Internet of things (h-IoT), mobile health and virtualisation (Gleiss *et al.*, 2021). Specifically, the h-IoT includes a diverse set of technologies proposed to improve disease management, remote monitoring of patients, fitness and wellbeing monitoring, health-related research, medication dispensation and

HealthCloud funded the research to understand their h-IoT ecosystem but had no influence during its design nor on the outcome. The authors declare no conflict of interest.

testing of treatments (Mittelstadt, 2017). These digital health technologies include a mobile sensor that captures physiological and activity-related signals, which are then processed and presented through an accompanying app (Kerr *et al.*, 2019).

Integration of devices and apps allows communication of user data across multiple platforms, mostly through an Application Programming Interface (API) or a software developer kit (Grundy *et al.*, 2017; Henriksen *et al.*, 2018). The integration of apps allows for data linkage to create a unified view of the user, incorporating data from a variety of spheres of life (Aitken *et al.*, 2016). Yet, ethical problems often arise and persist when data is shared across devices, apps and platforms and ultimately linked. h-IoT devices may generate personal health and activity data of extraordinary granularity and unparalleled scope, creating windows into the everyday, but private lives of individuals that are accessible by third parties for analysis purposes (Mittelstadt, 2017). Of particular concern are the protection of individual privacy, information manipulation, information leaks, loss of information and misuse (Grundy *et al.*, 2017; Mittelstadt, 2017; Kazlouski *et al.*, 2020). It is therefore prudent to evaluate the data sharing phenomena, for both research and commercial applications.

Fitness and nutrition apps are the most advanced in sharing data through APIs. Indeed, the integration of fitness apps has formed a network across which information may spread (Grundy *et al.*, 2017). Furthermore, fitness apps that function on a network of digital technologies have modified the exercise experience by extending it as a social experience through sharing of performance-related data with friends and strangers. The social dimension of fitness apps has dual implications for health and wellbeing, of which the negative impacts are regarded less than the positive effects (Whelan and Clohessy, 2021). In this article, we focus on one such potential negative implication, namely the concerns related to the privacy of users when data are shared across platforms.

Earlier work has been done to identify vulnerabilities and entities in the h-IoT ecosystem that facilitate data sharing, as well as the quantification of data volumes and data types that are shared within this ecosystem (Fereidooni *et al.*, 2017; Grundy *et al.*, 2017, 2019a; Kazlouski *et al.*, 2020; Tangari *et al.*, 2021). Although both Grundy *et al.* (2017) and Grundy *et al.* (2019a) were able to identify important entities and sharing pathways in subsystems of the h-IoT ecosystem using network analysis techniques, neither assessed the dynamic spreading capability of the networks. Therefore, with this article, we aim to offer a systems thinking perspective on the ongoing data sharing phenomenon, by analysing the hypothetical information spreading capability of a fitness apps network in terms of its structure and the centrality (positioning of influence) of its nodes (the apps). The systems thinking perspective conveys that the behaviour of a system (in this case, the information spread) is produced by the structure of the system and the interactions between the elements that make up the system (Maani and Cavana, 2007).

We contribute to the study of the data sharing phenomenon amongst fitness apps aimed at the consumer by quantifying the earlier descriptions of the potential spread of information using network methods. We chose network methods because it offers a succinct way to study the connection patterns (the behaviour) that result from the structure of a networked system, in which the elements are the nodes and the edges formed between the nodes (Newman, 2010). More specifically, at the network level, we assess the community structure of the fitness app network, as well as its assortativity (degree-correlation), since both these elements have been shown to influence information spread dynamics of other types of networks (Nekovee *et al.*, 2007; Weng *et al.*, 2013; Murakami *et al.*, 2017). At the node level, we assess spreading dynamics in terms of the starting node's degree (number of unique connections with other nodes) which also impacts the rate of spread (Moreno *et al.*, 2004). Furthermore, we assume the data shared amongst fitness apps are of the simple class of contagions (Weng *et al.*, 2013). Through the analysis, we highlight important aspects related to health ethics that may influence how individuals perceive their fitness and pose new questions.

Figure 1 indicates the focus of our study at the meso level of the h-IoT ecosystem. Individuals partake on the micro level by uploading and sharing their data on their chosen fitness app. Each app takes a position in the fitness app network at the meso level. The network of fitness apps is potentially capable of sharing data beyond what the individual originally intended when they upload their data.

The layout of the rest of the article is as follows: § 2 reviews the relevant literature, § 3 unpacks the method, § 4 contains the results from the network analysis and simulation, followed by the discussion in § 5. Supplementary material is available for definitions on network measures (§ S.1).

2. Literature review

2.1 The duality of data sharing

Technology in itself is neither good nor bad, but the purpose for which is applied may be beneficial or harmful to individuals or society (Pitt, 2019). Similarly, the data sharing phenomenon has positive advantages (understanding the subtle differences in individual behaviours, lifestyles and health, assisting healthcare providers to monitor and treat disease (Shah *et al.*, 2021), enhancing transparency between healthcare providers and patients to better chronic disease management (Dinh-Le *et al.*, 2019)) but may also cause harm (prejudicial treatment for insurance and employment, identity theft and blackmail (Slepchuk *et al.*, 2022), compromising optimal treatment through drug promotion when medicines-related app data are commercialised (Grundy *et al.*, 2019b), even undermining the free market (Pagliari, 2019).) Health-related data have travelled across boundaries between research, commercial and marketing contexts. Such cross-border sharing of data may potentially violate the expectations regarding the appropriate use of the data and erode public trust (Shah *et al.*, 2021).

To illustrate these points, Fereidooni *et al.* (2017) successfully manipulated data on fitness tracking apps by injecting fabricated values in proof-of-concept attacks on wearable device

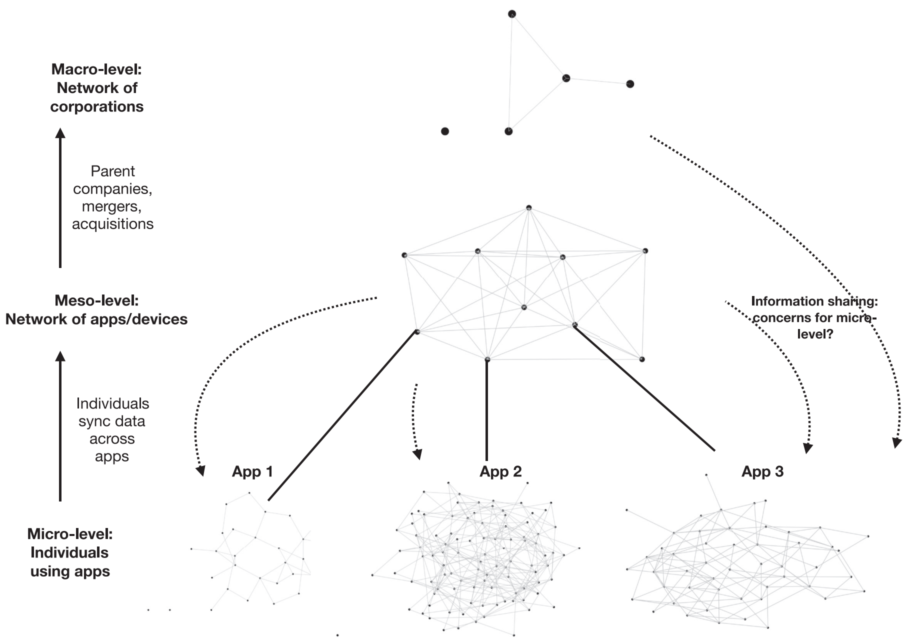


Figure 1.
Proposed levels of the h-IoT ecosystem

platforms, exemplifying the vulnerability of these apps. Kazlouski *et al.* (2020) showed that fitness mobile apps contacted unexpected third parties (location services, advertisers and analytic providers), continuing to state that individuals are likely donating private information to device manufacturers in exchange to use the app. Grundy *et al.* (2017) validated privacy and security concerns related to prominent health apps by tracing the potential flow of consumers' data through network analysis, in which they identified a core of 15 app families through which data may travel. Tangari *et al.* (2021) confirmed the presence of third parties that are traditionally outside of the health domain in the data traffic of health and fitness apps. Such third parties included targeted adverts, tracking, analytics, social networks, banking and games. In their sample, 88% of apps have the potential to collect and share user data. Grundy *et al.* (2019a) performed a network analysis on the third and fourth parties involved in medicines-related apps to understand how user data may be aggregated across the network. They provide proof that data sharing heightens the risk to privacy when data is aggregated by fourth parties. The research on the data sharing phenomenon has shown that data are possibly not shared for the benefit of the individual alone.

2.2 Network structure and propagation

Network analysis is a useful approach to finding patterns in the connections made between entities in real-life networks. Graph theory and statistical measures are combined to characterise the structure of a network. From this topological characterisation, the behaviour or evolving nature of a network may be better understood (Jalloul *et al.*, 2018). Network analysis has found applications in social networks, transportation systems, biology, economics, communication networks, Internet networks and bibliometrics to name a few (Jalloul *et al.*, 2018; Scott, 2017; Pool *et al.*, 2020; Fortunato and Hric, 2016). The structure of a network influences the dynamics of the contagion that diffuses across the network (Kuperman and Abramson, 2001; Zanette, 2002; Nekovee *et al.*, 2007; Weng *et al.*, 2013; Murakami *et al.*, 2017). Contagions can be classified as simple (infectious disease, information, rumours) or complex (behaviour, memes) (Weng *et al.*, 2013). Simple contagions only require a single contact to become "infected" (Centola, 2010), whereas common interest amongst community members (homophily) and social reinforcement are prerequisites for the spread of complex contagions (Weng *et al.*, 2013). The discussion here on network structure and diffusion dynamics is limited to simple contagions.

Special structures are formed in networks when nodes become organised into cohesive subgroups, referred to as communities, clusters, or modules. Such subgroups display higher connectivity (or edge density) within the subgroup than with nodes outside the community. Communities may develop some level of autonomy within the network and often have functional roles in networks (Fortunato and Hric, 2016). In social and organisational settings, communities can facilitate resource mobilisation, specifically attention, prestige, or information (Stoltenberg *et al.*, 2019).

Community structure influences the flow of information (including rumours) through a network in terms of propagation speed and total reach (Nekovee *et al.*, 2007; Weng *et al.*, 2013). A high clustering coefficient can inhibit rumour propagation (Jia *et al.*, 2020). Information may get trapped in communities because of the high internal cohesion, thereby limiting the global spread (Weng *et al.*, 2013). In small-world networks with local communities and a low number of long-range connections, it is possible that a rumour never leaves the neighbourhood of origin. On the other hand, as the randomness of the network is increased through more long-range connections, the rumour spreads through the network to reach a proportion of the population (Zanette, 2002). Strong community structure in hyperlinked networks creates densely connected areas that bind the surfer of the web to a particular area and by implication, the information this surfer attains is limited to that area (Stoltenberg *et al.*, 2019).

The initial rate of information spread was found to be faster in scale-free networks than in random networks with the rate of propagation augmented by assortative mixing in scale-free networks (Nekovee *et al.*, 2007). However, poor information spread was observed for both scale-free and random networks with high assortativity, in which the nodes with similar degrees have formed clusters with few connections between high- and low-degree communities. In addition, the average path length between nodes increased as assortativity reached the maximum value, attributed to the loss of shortcut edges due to the clustering effect (Murakami *et al.*, 2017).

3. Method

The initial network was formed with the original sample (§ 3.1) and whole-network and node centrality measures were calculated. This network was then pruned to maintain nodes with more than one connection, henceforth referred to as the pruned network, to find meaningful communities (§ 3.2) and to test information spread capability (§ 3.3). At this stage, to maintain parsimony but still find meaningful network structure, neither the direction of data sharing nor the actual fitness or health-related metrics that are shared or synced between apps were considered, only whether there is an existing connection between the apps. The network is therefore a simple, unweighted, undirected network.

Network analysis may be done with computational tools. The *igraph* library provides data types, functions and routines to implement graph algorithms for network analysis on a computer (CRAN, 2022). The *igraph* package in *R* was utilised as the computational tool in this research.

3.1 Sampling process

The snowball sampling approach was followed, whereby an initial selection of devices was made based on the content analysis from two Google searches (conducted in the Republic of South Africa), for (1) *best fitness tracking devices 2021* and (2) *best heart rate monitors 2021*. The top-4 or top-5 websites that the Google search returned were visited and downloaded into *.pdf* format. This number is based on diminishing click-through by users after the first four results (Jimenez *et al.*, 2019). The content of the websites was searched to identify potentially influential fitness devices or apps.

The identified devices were investigated on their respective websites, Google Play, or Apple's App store to find their partner or integration apps. An edge in the network is formed between nodes when there is a specific reference to another app(s) along keyword indicators of partners, "integrated with …", or a clear message regarding connecting to other apps through some synchronisation process. The second round of sampling consisted of finding the partner or integration apps of those apps that integrate with the initial sample following the same search process on their websites or the app stores. For example, round 1 (the content analysis) included Garmin, which has integrations with Strava, Google Fit and others (based on the information available on their website). Round two would follow to find Strava's, Google Fit's and other's integrations or connections through their respective websites. The edge list was then compiled after two rounds of sampling connections as mentioned between the apps on their websites or respective app stores. See Figure 2 for a visual aid to the sampling process. Technical data integration was not considered. This snowball approach reflects the reputational approach suggested in Scott (2017), where the initial devices identified through the content analysis are knowledgeable informants from which further information and connections are extracted.

For the network to represent independent entities as nodes, the devices or apps that belong to the same manufacturer or developer were grouped, for example, Garmin watches, Garmin heart rate monitors and the GarminConnect app were grouped to Garmin. Apple Health,

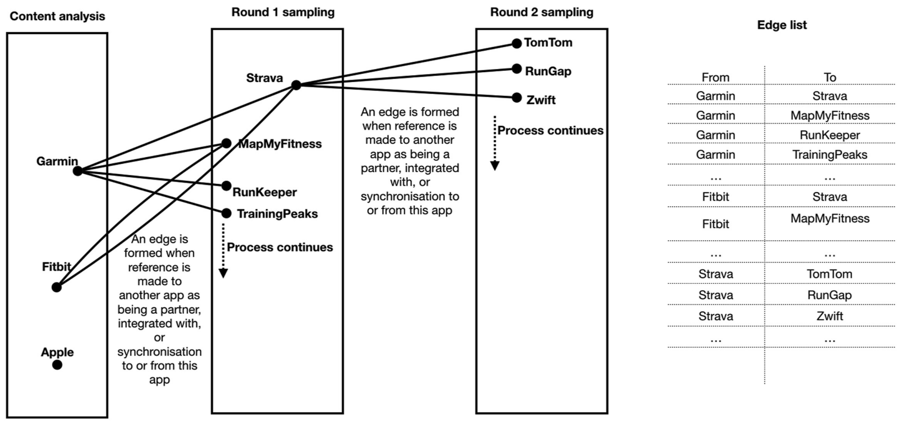


Figure 2.
The snowball sampling approach and edge formation between apps as nodes

iPhone, Apple Watch, HealthKit and so forth, were grouped to Apple. MapMyRide, MapMyRun and MapMyWalk were grouped in the MapMyFitness family.

3.2 Network communities

Communities in the network were assessed through node similarity (correlation, hierarchical clustering and k-medoids clustering) and community detection algorithms (walktrap, the Louvain method, label propagation, the Newman-Girvan method based on edge betweenness and the leading eigenvector algorithms). The Pearson correlation coefficient (ρ) is calculated for each node pair in the adjacency matrix A, into a similarity matrix, M. ($\rho \sim 1$ implies the nodes have similar connections, $\rho \approx 0$ implies no correlation and $\rho \approx -1$ implies dissimilar connections). The clustering algorithms take as input the distance matrix, D, which is equal to 1 - M. The dendrogram for the hierarchical clustering was divided at a height = 0.9 to deliver 12 clusters. The k-medoids algorithm in the *pam* function was applied with seven clusters (selected based on the elbow-test for the number of centroids at the lowest average silhouette index). The silhouette index is used to assess the quality of the k-medoid clustering result. ($S \approx 1$) indicates highly separated clustering, $S \approx -1$ indicates that overlapping of clusters occurred because some data points are closer to other clusters than their own (Aggerwal, 2015).) The modularity score is used to assess the quality of communities in the network (Fortunato, 2010) for each community detection algorithm and was collected with the *modularity* function. $Q \approx 1$ indicates strong community structure, $Q \approx 0$ indicates random structure (Newman and Girvan, 2004).

3.3 Information spread simulation

Here we simulate information spread as a simple contagion (Weng *et al.*, 2013), implementing the model in Alberto *et al.* (2020) in which they simulate gossip diffusion (considered a simple contagion) across a network of friends. We augmented their code to represent the information spread across the pruned network of fitness apps. Their algorithm sets one individual (selected at random) as the starting node by assigning their gossip indicator to true, all other nodes' gossip indicator is set to false. The algorithm goes through all the individuals in a random sequence and selects a random neighbour of the focal individual. If the selected neighbour is in possession of the gossip, the focal individual's gossip indicator updates to true to indicate that they now also possess the gossip. The algorithm iterates a set number of times, for each iteration re-assigning the gossip indicator of selected individuals based on the status of their neighbour. The gossip cannot be "unheard", that is a true status cannot change back to false.

In our augmentation, the information starts at a specific node and we repeat the simulation a number of times per selected node. This starting node represents the initial input of information or data synchronisation from the device when a user uploads their tracking data to a designated app. Time steps are discrete and information is spread to one randomly selected neighbour per time step. Three versions of the simulation on the pruned network were executed:

(1) The simulation was repeated 10 times for each node and endured for 140 time steps per simulation. In this simulation, each node is granted the opportunity to be the starting node.

(2) The manufacturer nodes (that is, the devices identified during the content analysis) were selected as starting nodes. The simulation was repeated 100 times for each of these nodes and endured 30 time steps.

(3) The top-10 connected nodes (based on degree centrality) were selected as starting nodes. The simulation ran 100 times for 30 time steps.

We generated a random network and a small-world network with the *erdos.renyi.game* and *sample_smallworld* functions in *igraph*, using as input the number of nodes, the edge density and the median degree from the pruned network. The re-wiring probability for the small-world network was set to 0.01 to generate a highly clustered network. The strengths of community characteristics (clustering coefficient, modularity score for the walktrap community detection) for the artificial networks are 0.61, 0.8 for the small-world network and 0.032, 0.26 for the random network. The spread of information was simulated across these two artificial networks as well, again using the same code from Alberto *et al.* (2020) with some adjustments. For all the network simulations we extracted the time step at which the proportion of nodes with the information has reached 100%.

4. Results

4.1 The content analysis

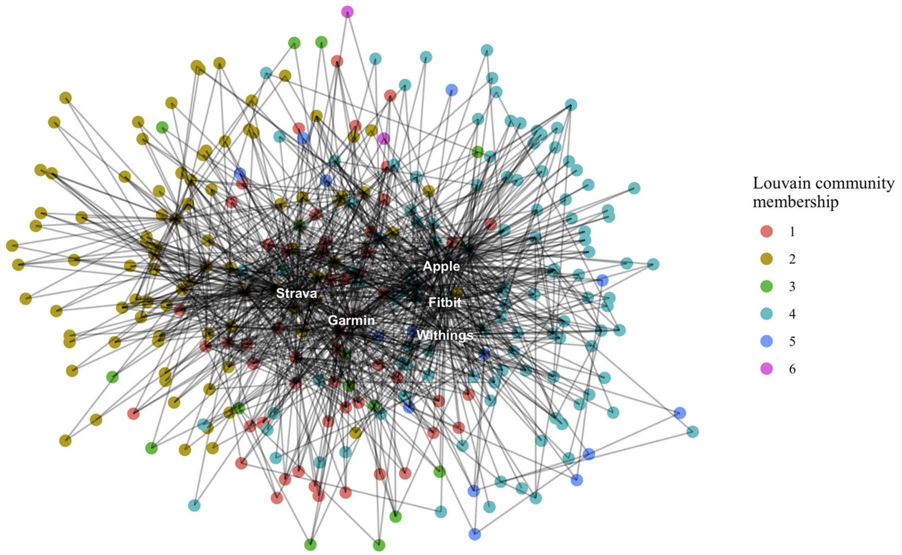
The supplementary material (§ S.1) lists the devices that featured consistently in the content analysis. Nearly all fitness trackers provide the number of steps (except Whoop), distances, calories, sleep tracking and optical heart rate (HR). Other metrics are assimilating into fitness trackers, namely: SpO_2 (7/13), electrocardiogram (ECG)-based HR and heart rate variability (HRV) (4/13), a-fibrillation indicators (3/13), fitness, stress or sleep scoring (10/13), breathing rate (4/13) and skin temperature (2/13). HRV requires high-quality interbeat data, preferably from ECG sensors and therefore remains elusive to fitness trackers. HR data from HR monitors are generated through either forearm photoplethysmography (PPG) sensors or ECG-based chest straps. HRV is available from nearly all ECG chest straps. Running dynamics, such as cadence and ground contact time may be recorded by some chest straps in addition to HR (Garmin, Miopod and Wahoo). VO_2-max data are used in fitness scoring, relating an individual's cardiovascular fitness to their age. Stress scores mostly incorporate HRV data to gauge the individual's stressed status, albeit algorithms differ in how this score is generated. The content analysis thus shows that highly personal and valuable medical data are being generated by fitness devices and apps.

4.2 Network structure and network-level measures

The final sample consisted of manufacturers (devices and sports apparel), fitness mobile apps, web-based applications, home automation apps, voice assistants, data platforms (or consolidators), retailer apps and software. The initial network (922 nodes, 1939 edges) presents a mixed picture in which we observed star-like structures as well as a core-periphery structure, which features a more densely connected core with a less dense periphery. This represents a scale-free network where the degree distribution seems to be following the power-law distribution (although not perfectly monotonically decreasing). The majority of nodes have less than five connections and a few have > 50 connections. Strava has the most connections (287), making it the most central node in the network based on degree.

The pruned network is produced by filtering the initial network for nodes with degree > 1, thereby maintaining 291 nodes connected through 1,308 edges. The core-periphery structure is maintained with some indication of star-like structure (Figure 3), as well as the power-law distribution (Figure 4). Strava remains the most connected node with 107 connections, indicating that it is an influential node in the network through which a large number of apps connect into the network.

Table 1 provides network-level measures for the initial network and the pruned network. The average path length decreased for the pruned network, whereas the density, clustering coefficient, mean degree and median degree increased for the pruned network. The assortativity is more negative for the pruned network than the initial network. (The average path length represents the number of data synchronisation steps a user must perform to shift their data across the network, that is to sync data from app 1, to app 2, to get it to app 3.) The density of 0.005 and the average path length of 3.2 for the initial network is consistent with the social network analysis of health and fitness apps by Grundy *et al.* (2017), who reported a density of 0.005 and 3.28 connections between two apps. The low density indicates very few of the theoretical potential connections in the network have indeed been realised, also reflected in the low clustering coefficient (that is, there is a low percentage of triadic closure). There are 631 nodes (68%) with only one degree, consistent with the median. It has been five years since the publication of Grundy *et al.* (2017), the low density that persists may be indicative that the network is not adding new connections amongst nodes despite potential changes in network size, or that the network is in constant flux during which new nodes are added but others leave the network. Smaller clusters away from the central nodes are therefore not forming.



Note(s): See online version for colours

Figure 3.
Network for health and fitness apps with degree > 1, coloured by their membership to communities based on the Louvain method. The membership is indicated by an integer number and does not carry any meaning itself

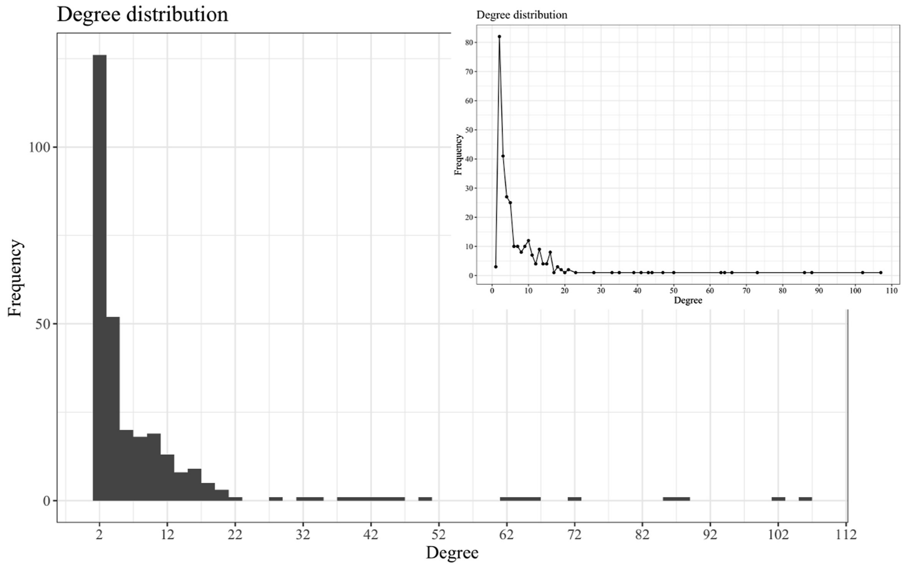


Figure 4. Degree distribution for health and fitness apps. Binwidth = 2 for the histogram. The insert projects the frequency for each degree

Table 1. Network-level measures

Measure	Initial network	Pruned network
Density	0.0046	0.031
Clustering coefficient	0.0443	0.105
Assortativity	−0.325	−0.398
Ave path length	3.2	2.56
Ave degree	4	9
St.dev degree	14.1	14.56
Median degree	1	4

The pruned network has a density of 0.031 and a path length of 2.56. There are 82 nodes (28%) with a degree of two and three nodes with a degree of one (due to structural changes in the network when edges were removed together with the one-degree nodes). We expected the increase in the density and clustering coefficient since the removal of nodes with one degree would result in a network with denser connections and a higher proportion of triadic closures. However, the density remains far below the theoretical maximum limit of 0.5 (Scott, 2017). The network remains disassortative, in which high-degree nodes connect with low-degree nodes. This is similar to the Internet-like networks in Vega-Oliveros *et al.* (2020). The disassortative topology indicates the presence of hubs through which most other low-degree nodes connect. The network is therefore heterogeneous in its topology, indicating that the connections did not form at random and that the entities of the subsystem of the h-IoT interact to form these connections.

4.3 Centrality measures

The pruned network is considered for the remaining analysis. Table S1 (supplementary material) presents the node-level centrality measures for the top-10 connected nodes. Manufacturers of fitness tracking devices make up seven of the top-10 nodes. This is indicative of how manufacturers may have expanded their product-service offering beyond their devices and data on fitness tracking. Strava is the most connected node, as well as the main broker node with a betweenness of 9,431, followed by Apple, Garmin, Fitbit and Withings. That the most connected node is also the main broker in the network not only highlights its influential role in the network but beckons the question of why this pure app node is more central than the manufacturing nodes that were identified during the content analysis. Strava is a social network platform dedicated to athletes on which they can track and analyse running and cycling performance, share their experiences with other athletes (friends) and form communities through clubs and challenges. In addition, Strava offers brands the opportunity to directly engage with athletes through sponsored challenges (Strava, 2021). In this way, Strava offers a resource to brands that is hard to come by, namely the attention of users (Stoltenberg *et al.*, 2019) (specifically, athletes who are more likely to engage with sporting brands).

4.4 Communities from similarity indices and clusters

Figure S1 (supplementary material) shows the right-tailed distribution of similarity as represented by the Pearson correlation coefficient, with most nodes' $\rho \approx 0$. Modularity scores from the community detection algorithms are on the lower bound for community structure that was formulated by Newman and Girvan (2004) ($0.3 \leq Q \leq 0.7$). The highest modularity is scored by the Louvain-method (six communities, 0.35), followed by the fast greedy algorithm (five communities, 0.33), walktrap (nine communities, 0.33), leading eigenvector (four communities, 0.31), Newman-Girvan (130 communities, 0.15) and label propagation (one community containing all the nodes, 0.0). The sizes of the respective communities for each algorithm are not evenly spread. All the algorithms except for label propagation formed one giant connected component consisting of most nodes (between 89 and 129) with smaller-sized clusters making up the rest of the communities. Cluster quality for the k-medoid algorithm is poor, with an average silhouette index of 0.11. There are 60 apps with a negative silhouette index, indicating the likelihood is reasonable that they belong to another cluster.

Products with the same core competency (for example the multisport watches: Garmin, Suunto, Polar, Coros) did not cluster together with hierarchical clustering, that is, they follow dissimilar connection patterns when compared with each other. The k-medoids clustering put together Garmin, Polar and Coros but Suunto was assigned to another cluster. All the community detection algorithms put these products together in the same cluster. That these comparable products form distinctive connections is good for the network and users since it is indicative of their preserved independence in the network; it also makes sense, when viewed from a competition point of view. To stand out in the market, comparable products form different connections to add to their product offering and so distinguish themselves from their competitors. The tendency of the network not to form strong disparate communities reflects the idea that the network does not rely on the autonomy of communities to function as a whole. Instead, we propose that devices and apps are forming connections with other apps to enhance service offerings for their users and cohesive communities within this network may hinder such diversifications.

4.5 Information spread

The time steps to reach 100% information spread for the three networks present with right-tailed distributions (see Figure 5). Median time steps to fully diffuse the information for each network are 25, 15, 9, 8 and 10 for the small-world network, the fitness app network, the fitness app network with the manufacturers as starting nodes, the fitness app network with the top-10 nodes (degree centrality) as starting nodes and the random network respectively. Figure 6 shows an inverse relationship between the starting nodes' degrees and the median time steps to reach full dissemination. From this simulation, we can envisage the rate at which

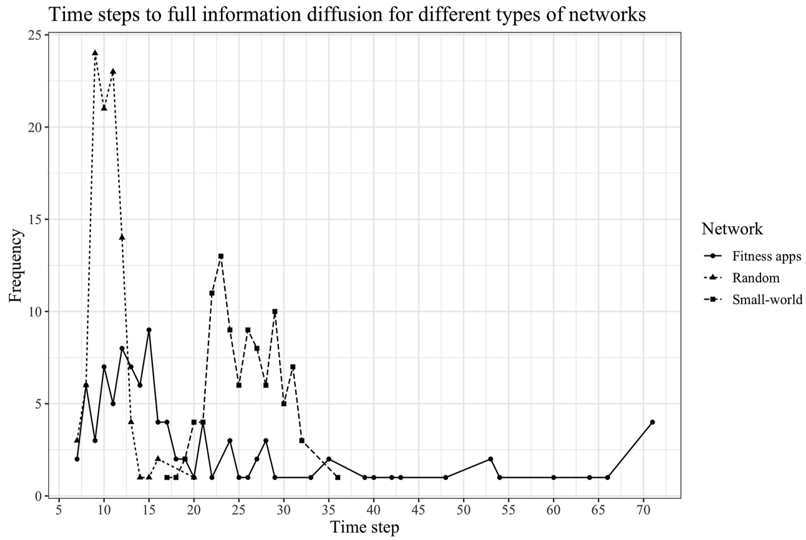


Figure 5.
Distribution of time steps for information spread

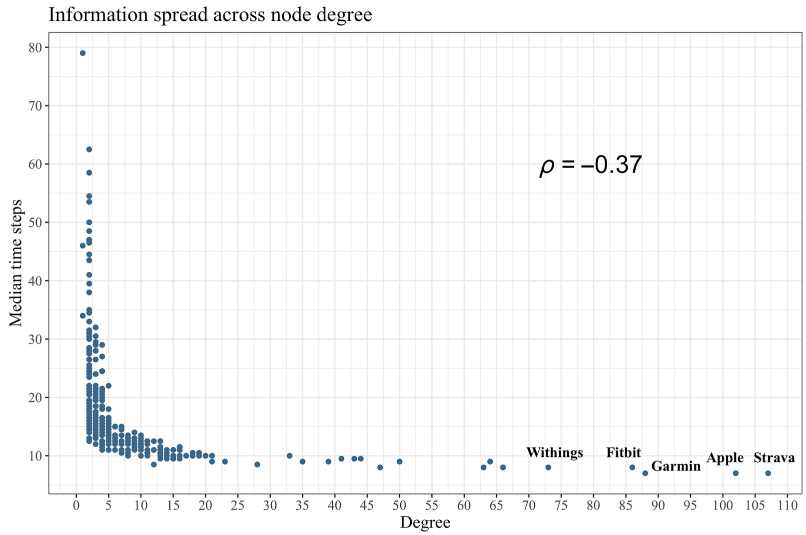


Figure 6.
Information spread across node degrees. The negative correlation (ρ) between degree and median time steps supports the idea that a higher degree starting node tends to disseminate information faster

health-related information may travel to nodes beyond what the user of the starting app originally had intended when they upload their data or provide other relevant personal information for the app to function. This rate of information diffusion is fastest for starting nodes with the highest degree, followed by the manufacturing nodes and differs from the artificial networks' rate of diffusion. The structure of the fitness app network, therefore, influences its diffusion capability, since there are no cohesive communities to retard the flow (such as those present in the small-world network).

5. Discussion

5.1 Communities and information spread

The weak modularity scores (closer to 0 than 1), the low silhouette index (close to 0), the correlation distribution around 0 and the low clustering coefficient are indicative that the network has not formed disparate communities. Instead, we observed a disassortative network with high-degree nodes connecting to low-degree nodes. The implication of such a community-poor network is seen in the pace of the simulated information spread across the network. The well-structured communities in the artificial small-world model with the low number of long-range connections slowed down the spread of information, whereas the lack of smaller, highly-connected subcomponents in the network coupled with the existence of a giant connected component likely accelerate the spread of information, further augmented by the hubs in the disassortative fitness app network. The pace of information spread is a function of the starting node's degree. We observed an inverse relationship between node degree and the time steps to complete the information spread. This is in agreement with Moreno *et al.* (2004) who showed that information (specifically, a rumour) propagates faster through the network when the initial node has a high degree of connectivity.

This has also been shown in other areas of study, e.g. Everett Rogers' diffusion of innovations and the effective dissemination of health-related messages through the social networks of highly-connected opinion leaders (Holliday *et al.*, 2016). In the case of the fitness app network, this dissemination of information may have dual effects. On one hand, individuals can intentionally move their data across the network in a few steps to the app of their choice after uploading to their device manufacturer's app. On the other hand, unwanted data sharing is possible even if the starting node has a low degree.

5.2 Concerns

The network extended beyond health and fitness applications to reach automated home appliances, such as the Philips Hue light bulbs, voice assistants (Siri, Amazon Alexa, Google Assistant) and retail reward systems (Walgreens, Dick's Sporting Goods). Medical apps (Dexcom, for continuous blood glucose monitoring, Qardio for blood pressure monitoring) connect to the health and fitness network through Apple, Fitabase, healow, FitnessSyncer, MyFitnessPal, Google Fit and others. Such connections are indicative of the emerging ubiquity of health and fitness concepts, driven by smart devices.

It seems that competing consolidation platforms are connected through communal affiliation with suppliers of activity data, similar to competitors of physical goods being connected when they source from the same supplier. Data consolidation platforms have addressed the need for health data consolidation, through integration and making it available on one platform. Perhaps a fitting analogue is that of the supermarket for data where businesses or consumers can "shop" for the data they require to address some need, be it operational or private health monitoring. Other products or service offerings may be built off this platform. Examples of extended service offerings are reward systems, health or fitness scores and engagement with users through chatbots which may allow for health information dissemination or collection of self-reported user data. However, the quality of the data is limited, because it is subject to being made available by individual devices or electronic sources and the veracity of the data. It may not be suited for monitoring purposes, since monitoring requires data to be collected in a consistent manner, which cannot be guaranteed in the case of consumer wearables due to varying algorithms and sensor qualities (Gay and Leijdekkers, 2015).

Voluntary, anticipated, but also involuntary and unanticipated, data sharing across health and fitness apps were highlighted in the network analysis of Grundy *et al.* (2017) in which they traced the potential flow of individuals' data through highly connected app families. The same applies to the analysis performed here. A risk to the network is a product of the network itself, namely its capability to share information. This sharing of private, health-related information allows comprehensive views to be built of the consumer, something that was not possible in previous decades. The network in this article has “made contact” so to speak with entities traditionally outside the health and fitness domains (for example, automated home appliances, retail reward systems and so forth). Through data consolidator platforms (examples in this network are Elcies, Human API and dacadoo) insurers and marketers can access personal health data in one place. In this way, the concerns of individuals who are reluctant to share their fitness tracking data with researchers due to fear that commercial third parties such as health and life insurers may gain access to it, have indeed materialised (Aitken *et al.*, 2016; Clarke *et al.*, 2021). It is important to maintain transparency with clients and users as to how the data are used for what purposes, who owns the data and who has access to it.

There might be top-down consequences from the information sharing across the network. When individuals realise how their data might travel across the network without them knowing it, even in anonymised form. The original intention of fitness apps is for personal monitoring at the individual level and not to share the data with other actors (for instance researchers, insurers, or even healthcare providers) (Clarke *et al.*, 2021). However, in light of the diffusion capacity of the fitness app network and the value that third parties may gain from a unified view of the consumer (Grundy *et al.*, 2017), we propose that the data sharing phenomenon moves away from the original purpose of wearables and their accompanying apps. Using data for purposes other than the original intention creates suspicion on the behalf of users and government alike – affecting the uptake of these devices and potential impacts that may be achieved (Clarke *et al.*, 2021). This creates any hitherto unanswered questions: Will data-sharing change users’ interaction with their devices? For instance, will individuals “hide” selected data or not use the device as intended to protect their fitness profile when viewed by insurers, reward programmes, or health service providers? How might information sharing impact the long-term adoption of wearable devices or other h-IoT-related devices? Long-term adoption of wearables is associated with sharing on social media (Friel *et al.*, 2021), but also with the decision not to engage with competitive step counting (Li *et al.*, 2020). This indicates that while some may indeed be motivated by public (or social) sharing of their data for the purpose of acknowledgement or competition, others may become demotivated when the data create a competitive milieu amongst users.

5.3 Implications

We have shown how fast data may travel through a subset of the network of fitness and health apps, given bidirectional flow. Such dissemination may benefit the user and society at large but may also cause harm, as discussed earlier in § 2. In light of the work done by Kuperman and Abramson (2001) who demonstrated that long-term oscillatory behaviour of epidemics is influenced by the community structure of the network, we now draw a parallel between data sharing and the challenges in maintaining and controlling the ongoing COVID-19 pandemic. We witnessed the spread of COVID-19 in waves across our world’s complex network structure, consisting of people and organisations. Yet, the interactions across this same network structure keep our economies operational. The public health challenge is therefore to maintain operational activities to spread economic benefits and at the same time limit the spread of the disease. Similarly, our work perhaps highlights that the core challenge for ethical data sharing is a multi-objective problem, since the structure of the network enables the duality of the purpose of data sharing: How do we maximise the benefits of fast data dissemination across the h-IoT, whilst minimising the risks associated with unwanted data sharing when both benefit and risk are achieved on the same network structure?

Future directions to address this question may involve a shift in how we approach the protection of privacy. Kasperbauer (2020) suggests three fundamental strategies (obfuscation of health data, penalties for misuse and improved transparency around with whom and for what purpose data are shared) when we consider the notion that once data are shared, we have given up control over our data and that the principle "privacy by design" undermines beneficial initiatives based on data sharing.

5.4 Limitations

This study is limited by the data that came available through the sampling process. Google applies country-specific strategies in their search algorithms, therefore the results of the content analysis may be limited to the relevant websites produced by the search conducted in the Republic of South Africa. We only had access to the information supplied by the websites, therefore the number of connections is limited to those that the website chose to reveal. Some apps and websites could not be reached during the network sampling process. Also, the information required to establish the direction of the edge is not provided uniformly on apps' websites nor the associated app store. The snowball approach has a higher likelihood to form a denser network than a random sampling approach, although the random sampling might produce an underestimation of the connection density (Scott, 2017). The network presented in this article represents a cross-sectional analysis of the network and we assumed that this cross-sectional analysis is representative of the full network of health and fitness apps. We approached the network as an undirected, unweighted network, but considering the directions of connections between nodes and the weights of edges may yield different insights.

Future directions for the network may include the application of information propagation theorems instead of a simulated experiment where information is shared between the nodes without a probability or rate measure. The intention to share information with selected apps is not incorporated into the simulation we applied here. Furthermore, looking for overlapping communities (Zubcsek *et al.*, 2014) and alternative ways to find the most influential nodes as starting nodes (Jia *et al.*, 2020; Vega-Oliveros *et al.*, 2020; Guo *et al.*, 2020) may also yield different information propagation patterns. Network resilience is another factor that can be explored to increase our knowledge of the network's structure and its influence on information spread.

6. Conclusion

This study extended the knowledge of the data-sharing phenomenon between fitness apps as a subset of the h-IoT by providing proof that the community structure, its disassortativity and the degree centrality of the starting node influence the information spread dynamics. Drawing upon the systems thinking perspective, we showed that the structure of the system (that is, the existing connections between apps) influences its behaviour (data sharing). However, data sharing itself is not the problem *per se*, but rather the uncontrolled and misappropriation thereof. The beneficiaries of our study are those researchers, policymakers, developers and lawmakers concerned with managing data sharing within the broader h-IoT, but also the individuals who wish to protect their privacy. We recommend a systems thinking perspective be taken to address the duality of data sharing. Other system thinking tools are available to better understand how the propagation of information may affect individuals and society. These tools include causal loop diagrams, agent-based and system dynamics simulations and machine learning, amongst others.

References

Aggerwal, C.C. (2015), *Data Mining: The Textbook*, Spinger, New York.

Aitken, M., de St Jorre, J., Pagliari, C., Jepson, R. and Cunningham-Burley, S. (2016), "Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies", *BMC Medical Ethics*, Vol. 17 No. 1, p. 73, doi: 10.1186/s12910-016-0153-x.

Alberto, A., Alex, M. and Marco, S. (2020), "Individual-based models of cultural evolution. A step-by-step guide using R", doi:10.31219/osf.io/32v6a, available at: https://acerbialberto.com/IBM-cultevo/ (accessed 19 November 2021).

Centola, D. (2010), "The spread of behavior in an online social network experiment", *Science*, Vol. 329 No. 5996, pp. 1194-1197, doi: 10.1126/science.1185231.

Clarke, H., Clark, S., Birkin, M., Iles-Smith, H., Glaser, A. and Morris, M. (2021), "Understanding barriers to novel data linkages: topic modeling of the results of the LifeInfo survey", *Journal of Medical Internet Research*, Vol. 23 No. 5, doi: 10.2196/24236.

CRAN (2022), "Package 'igraph'", available at: https://cran.r-project.org/web/packages/igraph/igraph.pdf (accessed 18 March 2022).

Csardi, G. (2019), "Package 'igraph', R", available at: https://igraph.org/r/doc/igraph.pdf (accessed 18 May 2021).

Dinh-Le, C., Chuang, R., Chokshi, S. and Mann, D. (2019), "Wearable health technology and electronic health record integration: scoping review and future directions", *JMIR mHealth and uHealth*, Vol. 7 No. 9, doi: 10.2196/12861.

Fereidooni, H., Frassetto, T., Miettinen, M., Sadeghi, A.R. and Conti, M. (2017), "Fitness trackers: fit for health but unfit for security and privacy", in O'Conner, L. (Ed.), *Proceedings - 2017 IEEE 2nd International Conference on Connected Health: Applications, Systems and Engineering Technologies*, CHASE, pp. 19-24, doi: 10.1109/CHASE.2017.54.

Fortunato, S. (2010), "Community detection in graphs", *Physics Reports*, Vol. 486 Nos 3-5, pp. 75-174, doi: 10.1016/j.physrep.2009.11.002.

Fortunato, S. and Hric, D. (2016), "Community detection in networks: a user guide", *Physics Reports*, Vol. 659, pp. 1-44, doi: 10.1016/j.physrep.2016.09.002.

Freeman, L. (1978), "Centrality in social networks conceptual clarification", *Social Networks*, Vol. 1 No. 3, pp. 215-239, doi: 10.1016/0378-8733(78)90021-7.

Friel, C.P., Cornelius, T. and Diaz, K.M. (2021), "Factors associated with long-term wearable physical activity monitor user engagement", *Translational Behavioral Medicine*, Vol. 11 No. 1, pp. 262-269, doi: 10.1093/tbm/ibz153.

Gay, V. and Leijdekkers, P. (2015), "Bringing health and fitness data together for connected health care: mobile apps as enablers of interoperability", *Journal of Medical Internet Research*, Vol. 17 No. 11, p. e260, doi: 10.2196/jmir.5094:10.2196/jmir.5094.

Gleiss, A., Kohlhagen, M. and Pousttchi, K. (2021), "An apple a day – how the platform economy impacts value creation in the healthcare market", *Electronic Markets*, Vol. 31, pp. 849-876, doi: 10.1007/s12525-021-00467-2.

Grundy, Q., Held, F.P. and Bero, L.A. (2017), "Tracing the potential flow of consumer data: a network analysis of prominent health and fitness apps", *Journal of Medical Internet Research*, Vol. 19 No. 6, p. e233, doi: 10.2196/jmir.7347.

Grundy, Q., Held, F., Continella, A., Bero, L., Holz, R. and Chiu, K. (2019a), "Data sharing practices of medicines related apps and the mobile ecosystem: traffic, content, and network analysis", *BMJ (Online)*, Vol. 364, doi: 10.1136/bmj.l920.

Grundy, Q., Chiu, K. and Bero, L. (2019b), "Commercialization of user data by developers of medicines-related apps: a content analysis", *Journal of General Internal Medicine*, Vol. 34 No. 12, pp. 2833-2841, doi: 10.1007/s11606-019-05214-0.

Guo, C., Yang, L., Chen, X., Chen, D., Gao, H. and Ma, J. (2020), "Influential nodes identification in complex networks via information entropy", *Entropy*, Vol. 22 No. 2, doi: 10.3390/e22020242.

Henriksen, A., Haugen-Mikalsen, M., Woldaregay, A., Muzny, M., Hartvigsen, G., Hopstock, L. and Grimsgaard, S. (2018), "Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables", *Journal of Medical Internet Research*, Vol. 20, p. e110, doi: 10.2196/jmir.9157.

Holliday, J., Audrey, S., Campbell, R. and Moore, L. (2016), "Identifying well-connected opinion leaders for informal health promotion: the example of the ASSIST smoking prevention program", *Health Communication*, Vol. 31 No. 8, pp. 946-953, doi: 10.1080/10410236.2015.1020264.

Iacobucci, D., Wasserman, S. and Faust, K. (1994), *Graphs and Matrices, Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge, pp. 92-166.

Jalloul, N., Poree, F., Viardot, G., Hostis, P.L. and Carrault, G. (2018), "Activity recognition using complex network analysis", *IEEE Journal Odf Biomedical and Health Informatics*, Vol. 22 No. 4, pp. 989-1000, doi: 10.1109/JBHI.2017.2762404.

Jia, J., Liu, X., Zhang, Y., Li, Z., Xu, Y. and Yan, J. (2020), "Rumor propagation controlling based on finding important nodes in complex network", *Journal of Industrial and Management Optimization*, Vol. 16 No. 5, pp. 2521-2529, doi: 10.3934/JIMO.2019067.

Jimenez, G., Lum, E. and Car, J. (2019), "Examining diabetes management apps recommended from a Google search: content analysis", *JMIR mHealth and uHealth*, Vol. 7 No. 1, p. e11848, doi: 10.2196/11848.

Kasperbauer, T. (2020), "Protecting health privacy even when privacy is lost", *Journal of Medical Ethics*, Vol. 46 No. 11, pp. 768-772, doi: 10.1136/medethics-2019-105880.

Kazlouski, A., Marchioro, T., Manifavas, H. and Markatos, E. (2020), "Do you know who is talking to your wearable smartband?", *Studies in Health Technology and Informatics*, Vol. 275, pp. 142-146, doi: 10.3233/SHTI200711.

Kerr, D., King, F. and Klonoff, D.C. (2019), "Digital health interventions for diabetes: everything to gain and nothing to lose", *Diabetes Spectrum*, Vol. 32 No. 3, pp. 226-230, doi: 10.2337/ds18-0085.

Kuperman, M. and Abramson, G. (2001), "Small world effect in an epidemiological model", *Physical Review Letters*, Vol. 86 No. 13, pp. 2909-2912, doi: 10.1103/PhysRevLett.86.2909.

Li, L., Peng, W., Kononova, A., Bowen, M. and Cotten, S.R. (2020), "Factors associated with older adults' long-term use of wearable activity trackers", *Telemedicine and E-Health*, Vol. 26 No. 6, pp. 769-775, doi: 10.1089/tmj.2019.0052.

Maani, K. and Cavana, R. (2007), *Systems Thinking, System Dynamics: Managing Change and Complexity*, Pearson, New Zealand.

Mittelstadt, B. (2017), "Ethics of the health-related internet of things: a narrative review", *Ethics and Information Technology*, Vol. 19, pp. 157-175, doi: 10.1007/s10676-017-9426-4.

Moreno, Y., Nekovee, M. and Pacheco, A. (2004), "Dynamics of rumor spreading in complex networks", *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, Vol. 69 No. 6 2, pp. 066130-1-066130-7, doi: 10.1103/PhysRevE.69.066130.

Murakami, M., Ishikura, S., Kominami, D., Shimokawa, T. and Murata, M. (2017), "Robustness and efficiency in interconnected networks with changes in network assortativity", *Applied Network Science*, Vol. 2 No. 1, doi: 10.1007/s41109-017-0025-4.

Nekovee, M., Moreno, Y., Bianconi, G. and Marsili, M. (2007), "Theory of rumour spreading in complex social networks", *Physica A: Statistical Mechanics and Its Applications*, Vol. 374 No. 1, pp. 457-470, doi: 10.1016/j.physa.2006.07.017.

Newman, M. (2010), *Networks: An Introduction*, Oxford University Press, Oxford.

Newman, M. and Girvan, M. (2004), "Finding and evaluating community structure in networks", *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, Vol. 69 No. 2 2, pp. 026113-1-026113-15, doi: 10.1103/PhysRevE.69.026113.

Pagliari, C. (2019), "Commercial health apps: in the user's interest?", *BMJ (Online)*, Vol. 364, doi: 10.1136/bmj.l1280.

Pitt, J. (2019), "Being human in the days/daze of big data [editorial]", *IEEE Technology and Society Magazine*, Vol. 38 No. 3, pp. 5-9, doi: 10.1109/MTS.2019.2930260.

Pool, J., Fatehi, F., Hassandoust, F. and Akhlaghpour, S. (2020), "Health data privacy: research fronts, hot topics and future directions", *Studies in Health Technology and Informatics*, Vol. 275, pp. 167-171, doi: 10.3233/SHTI200716.

Scott, J. (2017), *Social Network Analysis*, 4th ed., SAGE, London.

Shah, N., Viberg Johansson, J., Haraldsdóttir, E., Bentzen, H., Coy, S., Mascalzoni, D., Jónsdóttir, G. and Kaye, J. (2021), "Governing health data across changing contexts: a focus group study of citizen's views in England, Iceland, and Sweden", *International Journal of Medical Informatics*, Vol. 156, doi: 10.1016/j.ijmedinf.2021.104623.

Slepchuk, A., Milne, G. and Swani, K. (2022), "Overcoming privacy concerns in consumers' use of health information technologies: a justice framework", *Journal of Business Research*, Vol. 141, pp. 782-793, doi: 10.1016/j.jbusres.2021.11.073.

Sohn, I. (2017), "Small-world and scale-free network models for IoT systems", *Mobile Information Systems*, Vol. 2017, doi: 10.1155/2017/6752048.

Stoltenberg, D., Maier, D. and Waldherr, A. (2019), "Community detection in civil society online networks: theoretical guide and empirical assessment", *Social Networks*, Vol. 59, pp. 120-133, doi: 10.1016/j.socnet.2019.07.001.

Strava (2021), "Features for athletes, made by athletes", available at: https://www.strava.com/features (accessed 7 December 2021).

Tangari, G., Ikram, M., Ijaz, K., Kaafar, M. and Berkovsky, S. (2021), "Mobile health and privacy: cross sectional study", *The BMJ*, Vol. 373, doi: 10.1136/bmj.n1248.

Vega-Oliveros, D., da Fontoura Costa, L. and Rodrigues, F. (2020), "Influence maximization by rumor spreading on correlated networks through community identification", *Communications in Nonlinear Science and Numerical Simulation*, Vol. 83, doi: 10.1016/j.cnsns.2019.105094.

Wasserman, S. and Faust, K. (1994), *Structural Balance and Transitivity, Structural Analysis in the Social Sciences*, Cambridge University Press, pp. 220-248.

Weng, L., Menczer, F. and Ahn, Y.Y. (2013), "Virality prediction and community structure in social networks", *Scientific Reports*, Vol. 3, pp. 1-6, doi: 10.1038/srep02522.

Whelan, E. and Clohessy, T. (2021), "How the social dimension of fitness apps can enhance and undermine wellbeing: a dual model of passion perspective", *Information Technology and People*, Vol. 34 No. 1, pp. 68-92, doi: 10.1108/ITP-04-2019-0156.

Zanette, D. (2002), "Dynamics of rumor propagation on small-world networks", *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, Vol. 65 No. 4, p. 9, doi: 10.1103/PhysRevE.65.041908.

Zubcsek, P., Chowdhury, I. and Katona, Z. (2014), "Information communities: the network structure of communication", *Social Networks*, Vol. 38 No. 1, pp. 50-62, doi: 10.1016/j.socnet.2014.02.001.

Supplementary material

S.1 Definitions and network measures

A node in a network represents the entity that forms connections with other entities in the network, also referred to as actors or vertices. A node is represented as a point on a graph. Edges, ties, or connections refer to an existing relationship between any two nodes (Aggerwal, 2015). The discussion here applies to undirected networks. Although variations exist in how centrality metrics and network measures are calculated, this article follows the same mathematical expressions referred to by the *igraph* package.

Types of networks

Networks are often classified based on the distribution of the nodes' degrees. Degree is the total number of connections (or ties, or edges) that a node has with other nodes in the network (Freeman, 1978). In a random network, two nodes connect when a randomly generated number is larger than a given probability. All nodes, therefore, have an equal chance of being connected. The degree distribution is normal. In the small-world network model, a regular lattice network is transformed by re-wiring a number of connections to a randomly selected long-range node. In this way, the average path length (§ S.1) between nodes is reduced significantly. Both random and small-world networks, therefore, have a homogeneous topology. The degree distribution of real-world networks often follow the power-law distribution (heavy right-tailed), in which a small number of nodes have a very large degree and the majority of nodes have a small degree. These types of networks are referred to as scale-free and have a heterogeneous topology (Sohn, 2017).

Network measures

Let n be the number of nodes in a network and e be the number of edges. Density D_s in a network is the proportion of actual edges (e) to the total number of edges possible, in Equation (1) (Scott, 2017):

$$D_s = \frac{e}{n(n-1)/2} \tag{1}$$

The total number of potential connections any node can make is given as $n(n-1)$, that is a connection to every other node except themselves. The total possible or potential edges equals $n(n-1)/2$ since an edge between A and B is the same as between B and A. Density of networks is, therefore, a function of network size (number of nodes) and is limited by it, since time constrains the number of connections that can be meaningfully sustained by any node. When rewards start to decline and it becomes too costly to make new relations, agents (nodes) may decide to stop making new connections. The density of a network declines as its size increases, with an estimated mathematical maximum limit of 0.5 (Scott, 2017).

A path is the sequence of edges (connections) between two nodes in the network without repetition; that is, how to get from node A across to node B through the connections that have been formed between nodes. The path length represents the distance between two nodes, or how many steps must be taken to reach node B from node A, by counting the number of edges between two nodes, without repetition of edges or nodes (Iacobucci *et al.*, 1994). The clustering coefficient (also referred to as transitivity) measures the probability that adjacent nodes are connected (Csardi, 2019). A triad of nodes i, j, k is transitive (they achieved triadic closure) if the connections between all three have been realised, that is there is an edge between i, j, between j, k and between i, k (Wasserman and Faust, 1994). In real-world networks, triadic closure represents the tendency to form clusters. For instance, two individuals who have a friend in common are likely to become connected, since there is a higher likelihood that they have similar backgrounds and therefore there may be opportunities or reasons to interact. The clustering coefficient is related to this concept of triadic closure (Aggerwal, 2015). The clustering coefficient measures the ratio between the closed triads and all the triads in the network (Csardi, 2019).

Betweenness is the number of times a node v appears on the shortest paths between any other nodes s, t in the network, shown in Equation (2) (Freeman, 1978).

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{2}$$

where σ_{st} is the total number of shortest paths between nodes s, t that may or may not involve v. Betweenness measures the extent to which a node plays an intermediary role by acting as a broker, gatekeeper, or bridge between other nodes in the network. Broker nodes or gatekeepers therefore also exert some extent of control over the nodes they connect (Scott, 2017).

S.2 Results

The following device manufacturers and apps featured consistently across the websites and were included in round 1 of the sampling process for the network analysis, namely fitness trackers: Amazfit, Apple, Asus, Coros, Fitbit, Garmin, Oura, Polar, Samsung, Suunto, Whoop, Withings, Xiaomi and HR monitors: Garmin, Miopod, Myzone, Polar, Scosche, Wahoo, Suunto, 4iiii. Table S1 contains the centrality measures for the top-10 connected nodes.

Name	Degree	Betweenness
Strava	107	9,431
Apple	102	6,596
Garmin	88	4,288
Fitbit	87	3,879
Withings	73	4,300
Wahoo	66	3,953
Polar	64	1,922
Google Fit	63	2,545
TrainingPeaks	50	2,083
Suunto	47	1,080

Table S1. Centrality measures for top-10 connected nodes

Figure S1 contains the distribution of similarity (based on Pearson correlation) between apps.

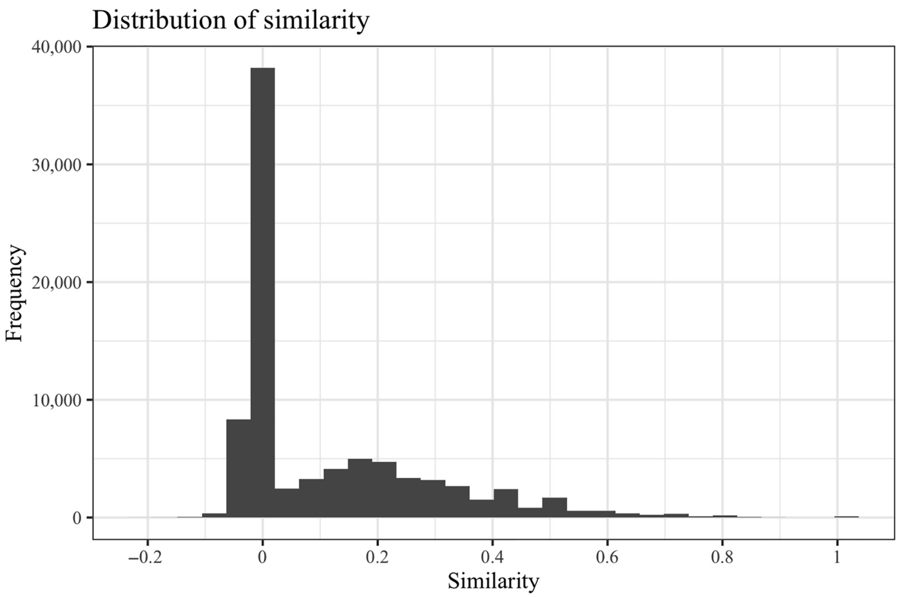


Figure S1. Distribution of similarity between nodes

Corresponding author
Euodia Vermeulen can be contacted at: euodiav@gmail.com