Distributed computing infrastructures have expanded over time from centralized clusters of computing nodes to decentralized wide area networks of heterogeneous devices. The latter can range from supercomputers to low-powered mobile devices. In particular, the rapid development of smartphone computing opened up new areas of research in distributed computing. With many-core CPUs and hundreds of gigabytes of storage, today’s smartphones are capable to perform complex tasks that only mainframes could run a couple of decades earlier. Augmented reality, image recognition, 3D rendering are only a few examples of CPU-intensive applications that can run on modern smartphones.

Not only are smartphones capable of running CPU-hungry tasks, but they are also able to support network intensive applications, like video streaming or social networking. This is possible due to the development and adoption of networking technologies such as WiFi, 4G and Bluetooth. While the first two are mostly used for making Internet connections, Bluetooth is mainly used for connecting smartphones to peripherals such as headphones or wearables due to its relatively short range. However, in recent years Bluetooth has evolved into a more general-purpose communication protocol, with ranges up to 200 meters. Other wireless communication technologies, such as Zigbee or LoRaWAN, have been motivated by the emergence of the Internet of Things, which demands wireless connections between co-located devices. This creates the potential for building distributed applications on top of ad-hoc mobile networks that leverage the spatial locality of devices to share and process data locally at lower latencies compared to a cloud-based infrastructure.

In this thesis we introduce and evaluate a distributed infrastructure in which mobile nodes organize and share data in a decentralized manner. In particular, we focus on ad-hoc networks of smartphones that self-organize into groups based on similarity. Our infrastructure consists of various components that are grouped on four
layers based on their computation and communication capabilities. This approach is based on the sensing-as-a-service model, which aims at providing structured access to sensor data. Similar to the sensing-as-a-service model, our infrastructure provides access to data in a network of mobile devices. However, our approach is different, as data is accessed in a decentralized way instead of using centralized access like in the sensing-as-a-service model.

On the bottom layer there are the sensors that produce raw data. These can be smartphone sensors, wearables or other sources of context information, such as user input or web APIs. These are accessed by specialized libraries that make the sensed data available to the applications that require it. Within our infrastructure, we use the SWAN library, as it is flexible and supports a wide range of sensors.

The second layer contains the sensor publishers, which are mobile applications that aggregate sensor data locally and distribute it through the cloud or within ad-hoc networks of co-located devices. While sharing data through the cloud has already been studied extensively, the notion of distributed ad-hoc sensing has been explored much less. Therefore, in Chapter 2 we analyse two methods of sharing data between co-located devices. The first method is based on the widely adopted Classic Bluetooth protocol, that provides larger bandwidth at the expense of higher energy usage. We optimize our method for various usage scenarios and test it in terms of performance and energy usage. The results show that it can adapt to frequent disconnections due to mobility and outperforms WiFi in terms of power usage.

The second method we propose leverages the Bluetooth Low Energy protocol, that was recently introduced to facilitate the transfer of small pieces of data between IoT devices. Compared to the first method, it incurs lower energy consumption and better usability for broadcasting sensor data. However, this comes at the expense of bandwidth, which is much lower compared to the Classic Bluetooth based method. We also show in our evaluations that using this method to receive sensor data from nearby devices is sometimes cheaper than using the onboard sensors.

On the third layer we find the service providers, which are cloud applications that collect sensor data from sensor publishers and process it in order to provide various services. To illustrate the utility of this layer, in Chapter 3 we introduce PeerMatcher, a service for grouping mobile nodes based on similarity. Any similarity metric can be used, as long as it can be expressed as a value for any pair of nodes. As opposed to existing clustering techniques, our service operates completely decentralized, meaning that nodes organize themselves into disjoint groups. We show that our service outperforms existing solutions in terms of speed and scalability, making it suitable for low-latency mobile applications.

To illustrate the utility of PeerMatcher, in Chapter 4 we describe a ridesharing system that uses PeerMatcher to cluster ridesharing users having similar routes in order to reduce cost and traffic. Our system fits well within the sensor data consumers layer of the sensing-as-a-service model, as it leverages a service provider in order to provide value to end-users. The novelty of our system lies in the fact that, like PeerMatcher, it is completely decentralized, therefore it does not require a
central operator to coordinate the matching process. The mobile nodes connect with each other directly and use PeerMatcher to discover and match with other nodes in a peer-to-peer fashion. We evaluate our system against an extensive dataset of New York taxi rides and show that it is able to cut the traffic and rides’ cost by more than a half.

The work presented in this thesis demonstrates the feasibility of a smartphone-based infrastructure for decentralized partnership formation, using existing network technologies. However, it also shows the need for technological improvements to create a more reliable infrastructure.