
WORKING PAPER

Oxford Future of Real Estate Initiative

Version 1.3 – 16th April 2020

PropTech: Turning real estate into a data-driven market?

Fabian Braesemann¹ | Andrew Baum¹

¹Future of Real Estate Initiative,
Saïd Business School, University of Oxford

Correspondence

Fabian Braesemann

Saïd Business School, University of Oxford,
OX1 1HP Oxford, United Kingdom

Email: fabian.braesemann@sbs.ox.ac.uk

Funding information

This work was supported by donors to the
Oxford Future of Real Estate Initiative.

Acknowledgement:

We would like to express sincere thanks to
Unissu and Crunchbase for the provision of
data.

Moreover, we would also like to thank the
founding donors to the Oxford Real Estate
Initiative (Forbes Elworthy of Craigmore
and The Centre for Studies in Property
Valuation and Management Trust) plus
Arcadis, BCLP, CBRE, EY, Grosvenor,
Nuveen, Redevco, The Crown Estate and
UBS for their financial support.

The real estate industry is traditionally a slow-moving asset class. The recent hype around real estate technology or 'PropTech' stands in stark contrast to this traditional view on real estate. It has been argued by PropTech entrepreneurs, tech evangelists and scholars that this 'digital disruption' of the industry leads to a digitalised global real estate market. These claims coincide with observations made in other markets that went through a process of digitalisation. Data-driven markets are often characterised by a winner-takes-all competition between firms that offer platform business models centrally focused on providing digital services for users, who 'pay' in providing more user data.

In this paper, we investigate whether PropTech is actually turning real estate into a data-driven market. The quantitative findings from an analysis of more than 7,000 PropTech firms reveal that such trends are at work in PropTech. PropTech is indeed an increasingly important, global phenomenon, with data analytics technologies at the core of the network of property technologies. In this core sector, most acquisitions between PropTech firms occurred.

The findings presented here are important for users and owners of real estate. In order to benefit from the efficiency gains associated with the digitalisation of the market, they need to become aware of the business value of data they are generating in buying, renting, or managing real estate.

1 | INTRODUCTION

Real estate, while being the world's largest asset class and central to the world economy (Saul et al., 2020), is traditionally a slow-moving sector. The industry has long been considered to be shaped by the importance of personal connections and a relative slow adoption of novel technologies (Fields, 2019b).

This narrative has recently been challenged by 'PropTech'. This movement¹ focuses on digital technologies and innovation; in stark contrast to traditional perspectives on real estate. PropTech entrepreneurs and tech evangelists

¹<https://www.unissu.com/proptech-resources/proptech-view-from-switzerland>

describe this movement as 'digital disruption' of the industry, leading to a digitalised global real estate market. This market is likely to be characterised by platforms and sharing, tokenised and fractionalised owning, working, financing and living. Instead of life-long investment decisions or leases agreed for many years, buying, owning or renting real estate could become a fluid process thanks, for example, to blockchain-based tokens. At the same time, smart homes, equipped with myriad sensors, communicate with the owner's smartphone to optimise energy efficiency and comfort.

Central to these innovations are digital data. At every step of the PropTech pipeline, data are either generated or processed to create value for users or owners of real estate. Consequently, it has been argued that real estate data become a tradeable commodity, and that data might actually become *the* central resource of the digitalised real estate market. The detachment of digital information about property from the materialised brick and mortar real estate business have led some to take a critical perspective on these developments. One concern is that the increasingly globalised and digitalised market might become more and more competitive, increasing the pressure to make the most profitable use of any piece of land, generating higher return on investment and potentially displacing less competitive users of real estate (Fields, 2019a).

At the same time, the importance of digital data as the lubricant of the PropTech machinery raises concerns about the potential platformisation (Kenney and Zysman, 2016) of the real estate industry. Over the past two decades, numerous markets have been datafied (Meyer-Schonberger and Cukier, 2012). While this process usually comes with substantial efficiency gains — think about the ease with which job applications can be submitted via online job boards or platforms such as LinkedIn — it often led to the concentration of market power in the hands of a few dominating firms. Datafied markets are shaped by winner-takes-all dynamics, because data involves material economics of scale. In recent years, digital giants entered numerous markets and captured the market shares for their competitive advantage. Recent developments in real estate indicate that similar winner-takes-all dynamics have started to take hold, and firms such as Amazon or Google have made first steps into the property market, while CoStar is already a dominant multi-billion dollar real estate data giant.

Against this background, it is important to understand the phenomenon of PropTech and its potential implications for the real estate industry as a whole. To investigate these trends, Oxford University's Saïd Business School has established the Future of Real Estate Initiative. Following up on the 2017 report 'PropTech 3.0' (Baum, 2017), the initiative recently released the follow-up report 'PropTech 2020: The Future of Real Estate' (Baum *et al.*, 2020), which describes the current state of development of the PropTech industry and its' most important technologies.

The study presented here complements the PropTech 2020 report by focusing more centrally on the role of data in PropTech. To do this, we collect and condense academic perspectives on datafication (Meyer-Schonberger and Cukier, 2012) and the PropTech phenomenon, from which we derive research hypotheses, which are then tested quantitatively. In analysing data about more than 7,000 PropTech businesses, we find that digital data are indeed at the core of the PropTech trend. Technologies that generate data or create value from digital data are central to the network of property technologies. Firms that apply such technologies have been most successful in obtaining external venture capital funding. Moreover, we find that PropTech is indeed becoming an increasingly global phenomenon, while being concentrated in high-income countries, and that winner-takes-all dynamics are most pronounced in those parts of the PropTech industry that deal directly with digital data. As in the broader tech industry, the U. S. and China, plus India, will likely use the size of their domestic markets to exploit economies of scale and develop across borders.

From these empirical findings we conclude that the important characteristics of datafied markets, such as issues related to data protection, fierce competition for market share, and the commodification of data become increasingly important in real estate. While this process comes with efficiency gains, it does not mean the benefits related to these gains are not likely to be distributed equally among those currently involved in the market. Users and owners of real estate need to become aware of the potential value of data that are generated about tenancy agreements, energy

consumption, or listing prices. Otherwise the benefits of datafication might be captured mostly by data-monopolists who offer complimentary apps and platforms to obtain real estate data for free from the users of their services.

2 | BACKGROUND

2.1 | Data-driven markets

As outlined in the PropTech 3.0 report, the application of technology in real estate is not a new phenomenon. With increasing computing power and the rise of the personal computer, the first wave of digital technologies shaping real estate business processes occurred in the 1980s and 1990s. This wave of digitisation in real estate coincided with globalisation trends in many sectors of the inter-connected world economy. For example, in 1983, American Airlines established its first international back office in Barbados (Manning *et al.*, 2017). Paperwork was flown from the U.S., handled and digitised in Barbados and sent back electronically via satellite to save 50 per cent labour costs. At the same time, manufacturing firms started to establish global production networks to optimise supply chains and cut costs (Yeung, 2018).

Such globalisation trends were made possible because of the ease with which information could be transmitted seamlessly thanks to improved information and communication technologies. However, in the 1980s, the costs of storing, processing, and transmitting digital data were still comparatively high. This limited the scope of extensive globalisation efforts to large corporations in manufacturing, insurance, or banking, and specifically to firms that could afford to set up back offices, local management, and extensive supply chains by themselves.

This trend started to change in the 2000s, when numerous web applications, in particular online platforms, started to enable internet-based business models. In the real estate industry, firms such as Zillow or Zoopla entered the market in that period, offering users vast choice and information to help make better buying decisions.² It is that period in which the foundations were laid for the digitalisation of the real estate industry, now subsumed under the term PropTech. Zoopla marks an example of a platform business model: the website offers its services to house buyers for free, but the users pay in the form of the data they provide. People who want to sell a property have an incentive to provide accurate information about the building or apartment (Boeing, 2020; Boeing *et al.*, 2020). Users signal their preferences in searching in specific neighbourhoods or applying filter options to get results that are relevant to them. During that process, the platform does not only observe the listing price – which usually serves as the market's condensed signal of all relevant information about a property – but gets a lot of useful information to estimate and predict market trends.

The case of Zoopla exemplifies another important characteristic of digital markets: people who want to sell or rent out an apartment have an incentive to use the platform with the largest user base. Those interested in buying a house are equally keen to get the largest possible choice from a platform with many houses being offered. Such 'network effects' (Easley and Kleinberg, 2010) are a major reason for competitive pressure in data-driven markets.

In pre-digital times, firms could offer a service to the local market which was largely detached from other markets due to geographical separation (Forman *et al.*, 2018). For example, a real estate agent in Oxford could serve the local market and use personal connections and information advantages to provide a sufficient choice of houses available to potential customers while being protected from competition with agents specialising in other cities. This changes with a digital platform business model: as users can switch from one digital platform to another without much effort, they tend to cluster on the platform that has the largest user-base and the economies of scale that follow, enabling

²Through the emergence of platforms such as Zoopla, Zillow, and Rightmove, now, 90 % of residential sales are generated by portals and only 5 % by estate agents' shop windows.

better technology and brand recognition through wider advertising. In such a business regime, the user decision to use platform A over platform B becomes a source of competitive advantage, and scale beats specialisation by location.

Because the marginal costs of providing one additional unit of an online service are essentially zero, this leads to a vicious competitive circle. As companies are keen to attract the largest possible user base, they will offer key services for free. However, providing the service, despite the near-zero marginal costs, is expensive: a server-infrastructure is needed to manage the web-traffic, data centres are required to store and provide relevant data, and a team of software engineers, web-designers, and customer support specialists are needed to maintain the smooth running of the website and provide the best UI/UX experience. Thus, platform-based businesses come with substantial fixed costs, for which significant funding is required. The fiercer the competition between platforms for market share, the more funding is needed to gain a large user base.

The combination of huge funding requirements to gain market share and the potential monopoly rents associated with market leadership fuel the race to attract huge amounts of venture capital to push digital platform business models.

Thus, we currently witness business dynamics that have information technology and data at their core. It is the information technology that enables digital business models in the first place. These business models collect, aggregate and process information. The processed information is essentially the service that adds value for users. While using the service, users provide another layer of valuable information to the firm, which can be the basis for novel products and services.

For example, the high-dimensional and granular data provided by sellers and buyers of houses on Zoopla provides a unique source of market information for the platform owners (Loukissas, 2017). It might have been that the platform did not see any immediate value from the user provided data. However, over time, with more sophisticated information technologies becoming available, novel applications could generate value from the user data. For instance, the listings information and the historical user search queries could be used to train automated valuation models (AVMs), which apply the latest statistical and machine learning models (Brynjolfsson and Mitchell, 2017). A platform such as Zoopla could then decide to either sell the data to other firms that offer AVMs or it could decide to provide such models to its own users, in order to attract an even larger user base.³ If the platform, in such a situation, has a larger database of historical user information than its competitors, it is likely that the platform's AVM will perform better than other valuation models on the market. This is because the model was trained on a bigger database, which captures a larger area of the data space describing the different house characteristics and user preferences. It is in such a situation that the real business value of data becomes apparent: data provides feedback to make statistical models more accurate (Mayer-Schönberger and Ramge, 2018). As accuracy is what users expect from a statistical model, they will use the service that provides the highest accuracy.

Another example: why would anyone want to use an online translating tool that provides poor results, if Google Translate performs so much better? In deciding to use Google's service because of its superior performance, the users provide additional feedback – in other words, data – to Google, thus helping to train the engine and improve the service even more. This leads to an increasing performance gap between the market leader's service and that provided by the competition. The result is a natural monopoly.

The key feature here is the non-rivalry of data as a good. Data do not get consumed by use (Ciuriak, 2018; Sadowski, 2019). Data can be used for different purposes in all imaginable combinations with other data sources and without losing value. Instead, because of the economies of scale just described, more data actually make models perform bet-

³In practice, Zoopla develops and sells AVMs through its subsidiary, Hometrack, whose clients include mortgage lenders. The U. S. company Zillow provides 'Zillow instant offers', enabling them to now invest directly into residential property as a company. This business, called 'iBuyer' accounted for half of the company's revenue in 2019 (Solomont, 2020).

ter, which usually attracts more users, and thus more data. This is another reason why boundaries between industries in digitalised markets have become blurred: firms that have gained a data-based competitive advantage in one market can easily transfer the data to another market to gain a competitive edge there.

For example, with Google Streetview, Google Earth, and Google Maps, Google has become *the* standard in navigation. As the firm offers all of its services for free, users are happily providing their data in return (Gordon, 2007; Lee, 2010). This data source becomes a critical source of competitive advantage in a different market: autonomous vehicles. Google's self-driving cars do not perform better because the company understands car-making better than Mercedes Benz or Volkswagen, but because Google's algorithms are trained on an massive amount of traffic-related data. Thus, in the digitalised business regime, data-rich firms from completely different industries could enter a market and use the data from 'their' home domain to get an advantage in the target market.

At the same time, the ease of information processing, open-source communities, the free online tools provided by the digital giants, and further accelerated globalisation due to common technical (software compatibility), business (international payments, International Financial Reporting Standards), and cultural (English language; Silicon-Valley like entrepreneurship) standards allows newcomers from all over the world to innovate novel services. The start-ups aim to digitally disrupt markets and to become global leaders in their market niche. The digital start-up hype we have witnessed over the past decade has in this way further fuelled the digitalisation and datafication of numerous sectors of the economy and it has also started to shape real estate. We will now turn to academic perspectives on the globalised and digitalised real estate market.

2.2 | The datafication of real estate

Digital data are now the driver of numerous applications in digital real estate. As the PropTech 2020 report describes in detail, there are many different technologies being applied in real estate, with three main areas of innovation.

Digital payments, alternative investment platforms and tokenised assets form key elements of *FinTech* innovations. These are relevant for real estate, not only as a way to invest, but also for payments and tenancy management, as they all help to reduce transaction costs (Forman *et al.*, 2018). Elements of the *Sharing Economy* (Sundararajan, 2016) allow for a better use of space. The best known examples, AirBnb and WeWork, have already substantially disrupted the hospitality and office markets with their concepts rooted in the sharing economy. Technologies that translate physical processes around real estate into digital data, i. e. sensors, cameras etc., form the core of the *Smart Buildings* trend in PropTech.

In all these elements of PropTech, data play a pivotal role. According to Porter (Porter, 2019, p. 575) PropTech as a phenomenon has co-emerged with the rise of Big Data (De Mauro *et al.*, 2018). The assemblage of 'technologies, platforms, apps and Big Data speed[s] up existing processes' and it increases the 'geographical reach' of real estate solutions. Indeed, it is the global scope of PropTech that constitutes a distinguishing feature from the local business regime that was traditionally a characteristic of real estate: Rogers (2017) and Sassen (2012) state that data and technologies have helped the development of globalised real estate investments (Rogers, 2017; Rogers and Koh, 2017; Sassen, 2012). Consequently the current phase of development marks 'the beginning of a digitally driven, global expansion' of real estate (Rogers, 2016, p. 24). In today's PropTech world, the real estate industry globally connects 'real estate agents, property developers, financial advisers, loan brokers, foreign investment lawyers, and IT professionals'. While space is not yet transportable, digital technologies (such as virtual reality) allow real estate platforms to detach information and capital flows from the materiality of real estate. Because of these trends it is possible for the PropTech industry to have a global scope.

Besides the increasing global scope, it is the commodification of data and the platformisation (Kenney and Zys-

man, 2016) that have been most intensely discussed by scholars. Keen (2015) describes the novel business regime as 'Googlenomics' (Keen, 2015): companies offer their software tools for free and get big data in return. Consequently, apartments and offices turn into big 'data factories' (Burrows and Savage, 2014). Tenancy platforms and smart devices transmit digital data, which are increasingly 'governing relationships, capital flows, and behaviours' (Porter, 2019). These tools offer 'opportunities for the production of data itself', which become a financial asset (Landau-Ward and Porter, 2019). While the improved matching and transactions are considered positively by the PropTech industry, Landau-Ward and Porter (2019) consider three aspects to be key to understanding the implications of PropTech (Landau-Ward and Porter, 2019): '*PropTech increases the sheer amount of recorded information, about land, housing, and property. Second, data digitisation has specific effects, such as the emergence of digital data as assets with value in and of themselves*' and '*PropTech brings new actors, products, and services into housing and real estate sectors*'. They conclude that a greater understanding of PropTech is vital to sharpen the legal and policy response to the emerging governance of digitalisation (p.582).

Shaw (2018) describes PropTech as 'Platform real estate' (Shaw, 2018). In applying diverse digital technologies, PropTech connects different actors in the real estate market. Key drivers of PropTech, which crucially depends on collecting and processing user data, are network effects, interoperability and standardisation, which Shaw predicts will become more important in real estate. As a consequence, Shaw predicts a fierce competition for market share ('Platform Wars').

In general, three themes emerge from the literature about PropTech. These are, first, the globalisation of real estate; second, the pivotal role of data; and third, the platformisation, that is, the competition for market share. In the next section, we translate these perspectives on PropTech into quantifiable hypotheses, which we will then investigate with a unique combination of data from two large online platforms that provide business information about start-ups and PropTech firms.

3 | HYPOTHESES

Based on the contributions reviewed in the previous section, we hypothesise that PropTech is a global phenomenon, mainly focused around the datafication of real estate and shaped by increasing winner-takes-all dynamics. At the core of PropTech is the detachment of the physical process of renting or selling an apartment or office building from the numerous information processing steps that are usually involved. This detachment allows globally oriented firms to offer digital services. Consequently, we can develop a working hypothesis:

Working hypothesis: PropTech, a global phenomenon, is the process of datafication of real estate. The economics of data lead to fierce competitive dynamics for market shares.

The working hypothesis is operationalised by three hypotheses. Many PropTech firms offer global solutions, and the potential for digital real estate processes is great in all parts of the world. Nonetheless, the development of PropTech industries is likely to be concentrated in the most urbanised, technologically leading regions of the world:

H1: PropTech is a global phenomenon - its distribution can be explained by macro-economic factors.

The contributions reviewed above state that data are at the core of PropTech. Instead of just improving real estate processes, data become a commodity by itself. Accordingly, we assume that data and technologies that generate

business value from data are at the core of PropTech:

H2: Technologies that generate value from data are essential in PropTech.

In datafied markets, such as marketing or human resources, winner-takes-all competitive dynamics have usually followed the increasing value of data-driven business models. Consequently, we would assume that there should be more acquisitions in those parts of the PropTech market which deal with digital data (Liggio and Haggerty, 2019):

H3: There are more acquisitions in the PropTech sub-sector focused on data than in other parts of the market.

4 | DATA AND METHODS

Before presenting the quantitative results, we briefly describe the data sources (Crunchbase and Unissu) and the methodologies involved.

4.1 | Data sources

Crunchbase is an online platform providing information on private and public companies, with a focus on early-stage firms. In particular, the website lists investments and funding information, the founding members of start-ups and a brief description of a firm's business model and 'tags' to describe the technologies used by a firm and the sectors a firm is operating in. Crunchbase currently lists more than 900,000 companies.

Unissu is a website that lists detailed information about more than 7,000 PropTech firms. In contrast to Crunchbase, which covers the complete universe of start-ups in all sectors of the global economy, unissu aims to inform the real estate industry. While it lists much fewer firms than Crunchbase, it provides more details about each individual PropTech company and the tags used to describe a firm's business model. Technologies are focused on real estate and PropTech. Unissu uses Crunchbase's company identifiers to match both platforms' information in order to provide funding information about each firm. Thus, the data from both platforms complement each other: Crunchbase provides figures on funding and acquisitions for the whole start-up universe, while Unissu provides more detailed information about PropTech firms, their technologies used and the real estate business context within which a firm operates.

In this report, we combine data from both platforms to generate the largest, most comprehensive dataset available to investigate the PropTech industry. While this dataset is much larger than any previously utilised set of firm data on PropTech, the results hinge on the definition applied to identify PropTech firms. At this point, we need to emphasise that our findings are limited by the PropTech definition used by Unissu. The platform defines a firm to be a PropTech company if it is '*primarily focused on supplying product(s) which apply to one or more of the lifecycle stages of any property asset, anywhere in the world*'. The lifecycle stages used are *Build, Buy, Sell, Rent/Lease, Move, and Manage*. While this categorisation covers many processes in the real estate industry, Unissu does not define firms active in the 'sharing economy' as PropTech firms, as the platform considers such firms as mere users of real estate. Thus, important companies like Airbnb and WeWork are not included in their data set, which has, for example, substantial implications on assessment of the total funding amount directed into PropTech.

4.2 | Creating the technology network

Besides descriptive statistics, we present a network of property technologies. In constructing the network from the Unissu technology data, we proceed as follows. Each firm is assigned one or more technology tags. A technology is represented as a binary vector containing a one if technology tag A (for example 'Big Data') is present and zero otherwise. On this dataset, we applied association rule learning Agrawal *et al.* (1993) to construct a network at yearly intervals. The association rule concept *lift* is used, as it provides a balanced measure of proximity between two technologies (tags). Formally, the lift between two technologies A and B is their joint occurrence probability divided by the technologies' unconditional probabilities:

$$\text{lift}_{A,B} = \frac{P(E_A \cap E_B)}{P(E_A)P(E_B)},$$

where E_A and E_B describe a firm using technology A and B , respectively. A lift > 1 implies that two technologies tend to occur together. Accordingly, this is the threshold for a link to be established between two technologies (nodes) in the network.

In the resulting yearly networks, we calculate the *normalised betweenness centrality*⁴ of the individual technologies as a measure of their importance. For more details of the methodology, see (Braesemann, 2019).

5 | RESULTS

In this section, we first present descriptive results on the growth of the PropTech market in order to set the context and provide empirical evidence for the claims about the increasing relevance of PropTech within real estate. Then, we present quantitative results on the three elements of the PropTech industry that were discussed above: the global scope of PropTech, the centrality of data-related technologies in PropTech, and the competition for market share.

5.1 | The growth of PropTech

Figure 1 shows the number of firms founded, total funding, and funding per firm in (i) PropTech, (ii) other real estate firms, and (iii) firms active in finance from 1998 – 2018. The figure displays the results on a logarithmic scale. Each dot represents the raw observations, while the solid lines represent smoothed trend lines to highlight patterns in the time series. PropTech firms are those 7,000 companies classified by Unissu. Other real estate firms are those firms that have a 'real estate' tag in Crunchbase, but which are not included in the Unissu platform, and finance firms (representing the FinTech sector, which is the sector that is most often compared to PropTech) are those that have a 'finance' tag in Crunchbase.

The left panel of the figure compares the number of newly founded firms per year in the three sectors. The figure reveals three major stages in the development of the start-up sectors, which are surprisingly consistent across the sectors: a first phase of slight growth from 1998 to 2003/2004; a second phase of exponential growth from 2005 to 2014; and a stage of saturation that started in 2015.

Unsurprisingly, during the first period from 1998 to 2003, the number of newly founded PropTech firms (dark blue) is very low. Only 30 – 60 PropTech firms were founded in these early years of the observation period. In contrast,

⁴The Betweenness centrality is the number of these shortest paths that pass through a node. A node's betweenness centrality is divided by the average betweenness centrality of all nodes in that year to allow comparison between networks.

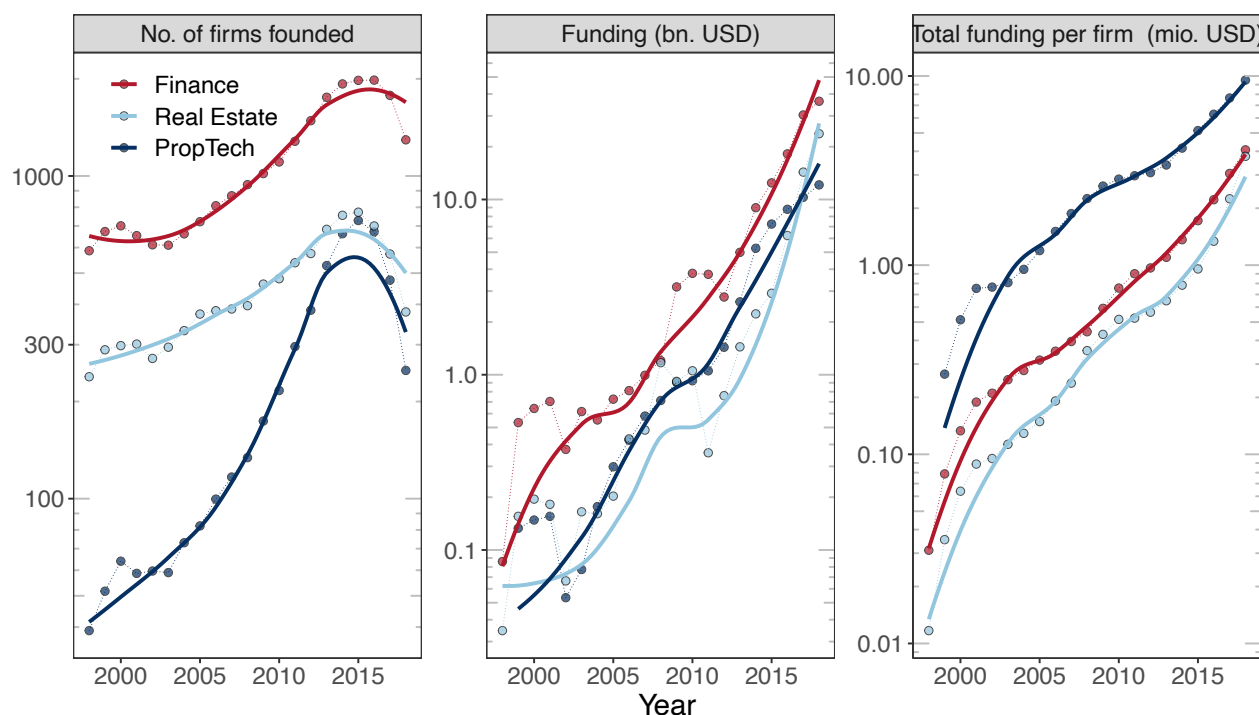


FIGURE 1 Number of PropTech firms and funding over time, compared to other real estate firms and firms in finance.

in the same period, there were 200 to 300 other real estate start-ups founded per year included in the Crunchbase dataset. In that period, around 500 – 600 finance firms were founded per year.

The second stage, which started around 2004 / 2005, was shaped by an exponential growth in the number of newly founded companies in all sectors. While the number of new start-ups was 60 in PropTech, 276 in other real estate, and 509 in finance in 2004, these numbers had increased to 661 in PropTech (+ 1,000 %), 708 in other real estate (+ 156 %), and 1698 in finance (+ 233 %) by 2015. In particular, the PropTech sector showed an exceptional growth rates, and by 2015 PropTech startup numbers had caught up with the number of non-PropTech real estate firms founded by 2015. The number of newly founded firms peaked in all sectors in that year, when a saturation phase began.

In contrast to the number of firms, the aggregate amount of funding (shown in the central panel of figure 1) grew exponentially over the complete observation period in all three sectors with comparable growth rates. While funding around 2000 differed substantially between finance (1.4 bn USD) on the one hand, and PropTech (379 million USD) and other real estate (371 million USD) on the other, both real estate sub-sectors would to some extent catch up with finance over time. In 2018, PropTech firms obtained 13 billion USD of venture capital funding, other real estate firms raised 20 billion USD, and finance start-ups obtained 50 billion USD funding in total.

While the total funding figures seem to vary widely between the sectors, these number are not adjusted for the differing sizes (defined by the number of businesses) of the three business sectors. Accordingly, we plot the average funding per firm in the right panel of figure 1. Again, the time trends in all three sectors look surprisingly consistent. In all industries, there was a period of steep growth before entering a long period of steady, but slightly slower growth. In contrast to the total funding, PropTech firms raised the highest amount of per-firm funding over the complete time span.

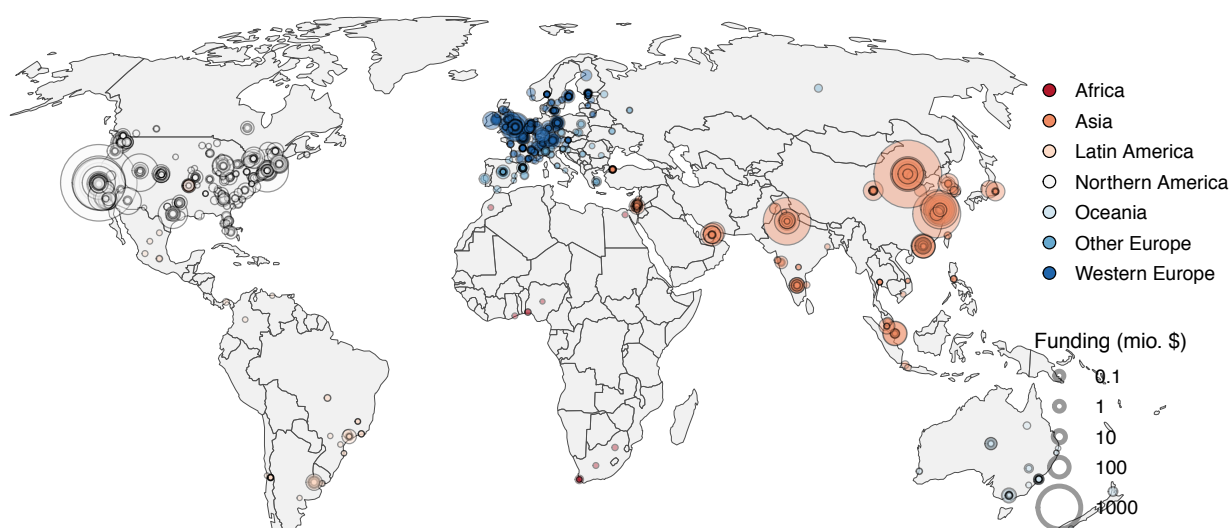
The difference with the other sectors becomes even larger over time. While finance firms in 2000 obtained, on average, 217,000 USD, other real estate firms got 99,000 USD, and PropTech pioneers 864,000 USD each. By 2018,

finance firms got 5.3 million USD funding on average, other real estate firms 3.5 million USD, and PropTech companies 10 million USD.

The main insight from figure 1 is that, while the start-up sector in total matured over the past two decades and eventually reached saturation in terms of the number of firms founded, funding has grown and is still growing exponentially. The PropTech sector has started as a niche within the larger real estate industry 20 years ago, but has now become a large industry by itself. All the time, PropTech firms have been able to secure higher than average funding, which was necessary to induce change in a slow-moving industry. Thus, we can conclude that PropTech is indeed becoming more important.

5.2 | The global geography of PropTech (hypothesis 1)

PropTech is a global phenomenon. As figure 2 shows, PropTech start-ups are distributed all over the world. Each dot represents a PropTech firm, while the size of the dot corresponds to the funding the firm has obtained. However, PropTech clearly clusters in specific regions. Hotspots of the PropTech industry are California, the US east coast, Western Europe (in particular the UK), and metropolitan areas in Asia (Delhi, Shanghai, Beijing, Seoul, Singapore). Compared to these places, most other regions of the world have much less developed PropTech sectors.



Map based on locations of 6,428 PropTech firms

FIGURE 2 The global distribution of PropTech firms.

While the figure reveals the geographical distribution of PropTech firms on a global scale, the results do not take the population of a country into account. A normalised statistic is displayed in figure 3. It shows the number of PropTech firms per one million population per country, grouped into larger geographical regions. The figure shows a sharp distinction between the economies of Western Europe, Oceania, North America with 6 to 9 PropTech companies per one million population on average on the one hand, and Global South (developing) countries on the other, which have 0.2 to 1.1 PropTech companies per one million population. These figures highlight the wide gap in the development of the PropTech industry between high- and low income countries.

In order to assess the development of national PropTech industries, it is important not just to consider the total number of firms, but also their capitalisation in terms of venture capital funding. As can be seen from figure 2, the

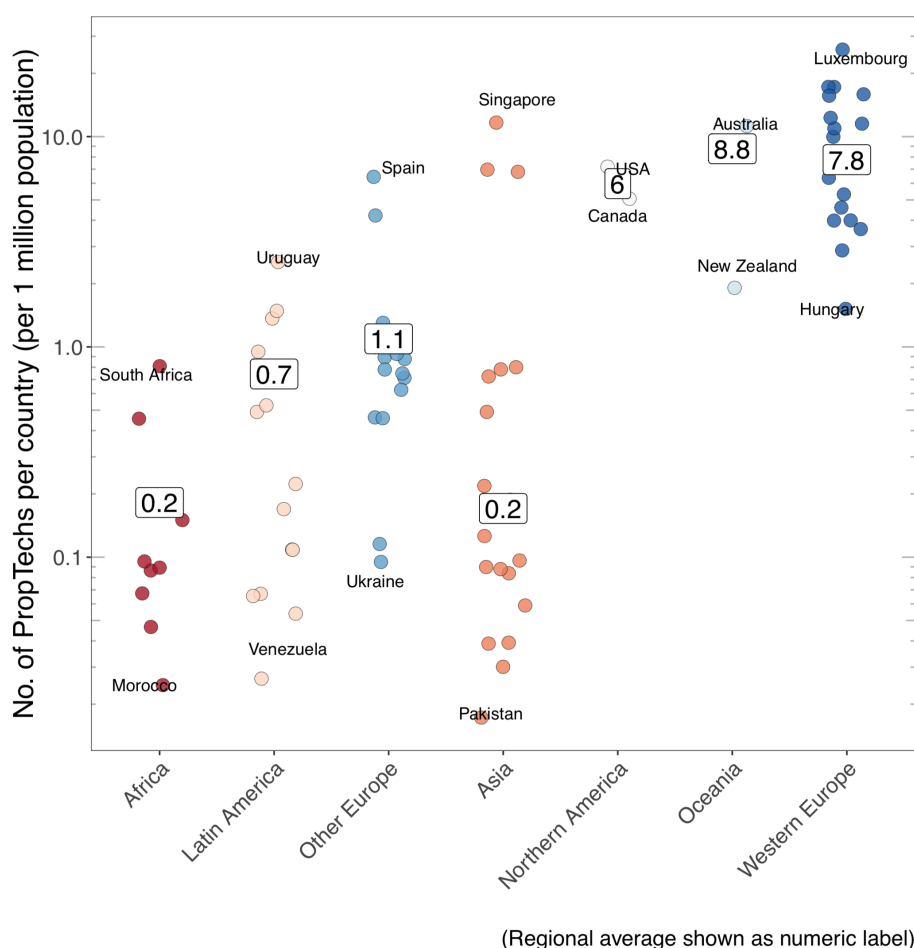
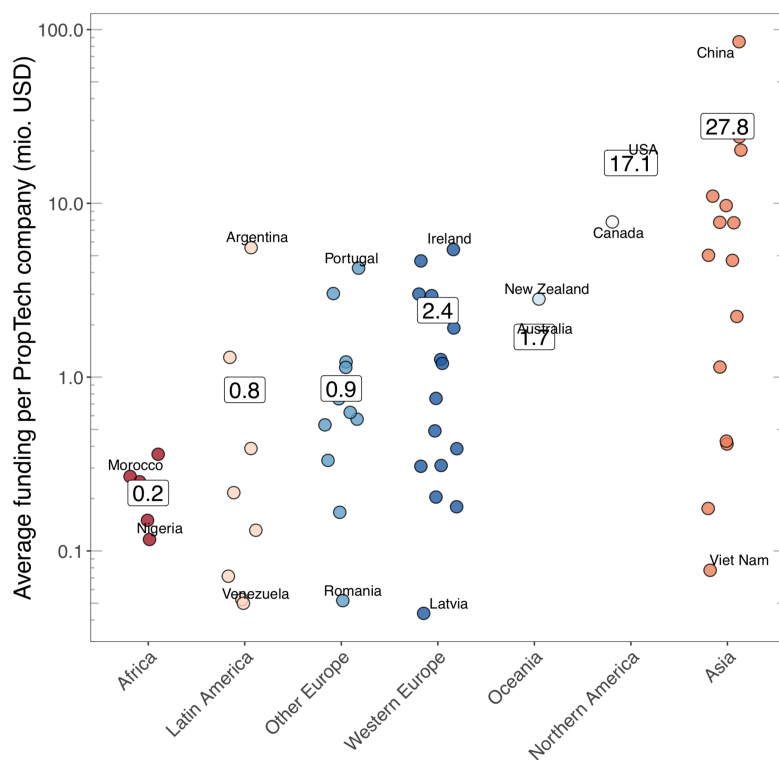


FIGURE 3 Number of PropTech firms per country (per 1 million population).

distribution of those firms that got the highest funding amounts (defined by the size of the circles) are even more spatially concentrated than PropTech companies in general. Most of the 'superstar' firms are located in San Francisco, Delhi, Beijing, and Shanghai.

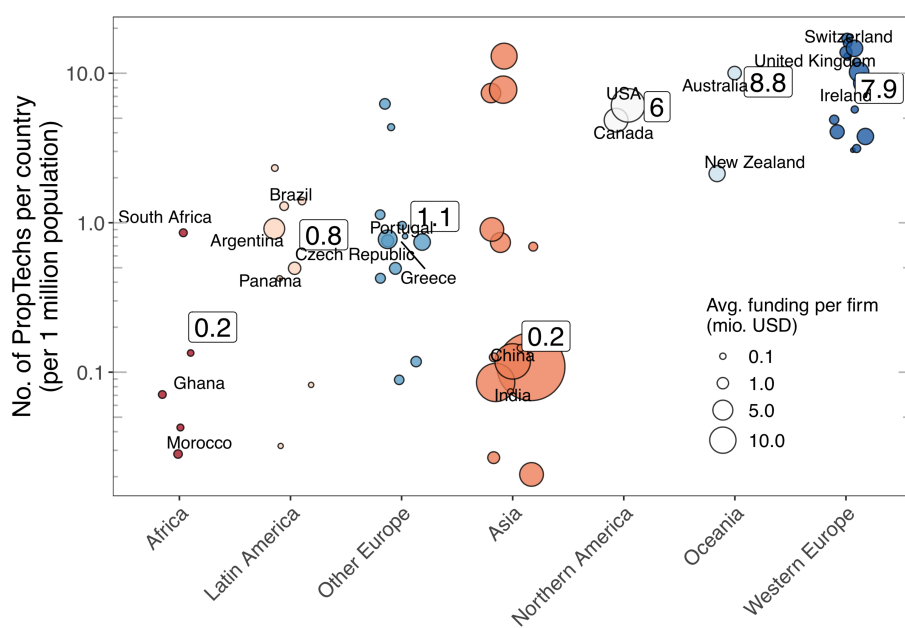
Figure 4 confirms these findings. It shows the average funding per PropTech firm in millions of USD per country. In contrast to figure 3, the average funding per firm shows more variation on a global level with less obvious macro-regional patterns. In particular the funding in Asia varies substantially, with Chinese PropTech companies accumulating 85 million USD of funding per firm versus a 28 million USD regional average (and Vietnam, for example, having obtained less than 100,000 USD funding per firm). There is much funding variation in the other macro-regions, as well. In Latin America, West and Other Europe, the average funding per firm varies by several orders of magnitude from around 100,000 USD funding per firm in Venezuela, Romania, and Latvia to more than 5 million USD funding in Argentina, Portugal, and Ireland. Funding in Africa is low overall (between 100,000 and 500,000 USD per firm) while it is high in the four countries in Oceania and North America that are included in the dataset.

More revealing than the number of PropTechs per population and funding per firm is the combined perspective presented in figure 5. In this figure, the vertical position represents the size of the national PropTech sectors in terms of number of firms per population, while the size of the dots represents the average funding. The countries with the highest average funding per region are highlighted. Interestingly, it is not the largest national PropTech sectors that show the highest average funding per region, but often economies that have intermediately sized sectors such as



(Regional average shown as numeric label)

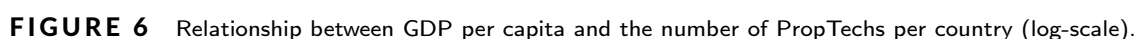
FIGURE 4 Average funding per PropTech company.



(Regional average shown as numeric label)

FIGURE 5 Number of PropTechs per population and average funding.

Switzerland, Ireland and Australia, all punching above their weight.



Despite these outliers, there are general trends relating the size of the national economy with the development of a domestic PropTech sector. As figure 6 shows, there is a clear relationship between the number of PropTechs per country (y-axis) and the GDP per capita (x-axis). High income countries have, on average, larger PropTech sectors. The size of the domestic PropTech sector, however, varies substantially and the strength of the local economy is not the only determinant, which can be seen from the wide spread in the figure.

Model 2 and model 3 show the relationship between macro-economic variables and total funding and funding per firm, respectively. In both cases, patents, as a measure of regional innovativeness, are positively associated with funding. Thus, it is the more innovative economies that are able to host the highest capitalised PropTech companies. While the size of the national PropTech industry seems to correlate mostly with the size of the local economy, total funding and funding per firm (as proxies of the international competitiveness of the PropTech industries) are more strongly driven by local innovativeness.

In summary, the findings presented in this section show that, while PropTech is a global phenomenon, the development of the national PropTech markets vary substantially, revealing global gaps in the number of firms and funding per country. While many high-income countries host relatively mature PropTech sectors with many firms, many low-

	PropTechs (1)	Funding (2)	Funding per firm (3)
Population (lhs-transformed)	0.93*** (0.17)	1.51 (0.94)	0.67 (0.83)
Patents (lhs-transformed)	-0.05 (0.11)	1.15* (0.58)	1.15** (0.52)
ICT-Exports (lhs-transformed)	0.28** (0.12)	0.64 (0.67)	0.47 (0.59)
GDP p. c. (lhs-transformed)	2.11*** (0.31)	3.84** (1.53)	2.10 (1.36)
Constant	-36.19*** (5.53)	-63.87** (27.99)	-32.60 (24.81)
Observations	69	69	69
R ²		0.67	0.58
Adjusted R ²		0.64	0.56

Note: * p<0.1; ** p<0.05; *** p<0.01

FIGURE 7 Determinants of the size and funding of national PropTech sectors.

income countries have huge development potential to establish firms that offer innovative property technologies to solve the most urgent problems of the local real estate markets. Given the high level of imbalances, it stands to reason that most of the smaller PropTech firms from the developing world will find it hard to compete against better funded competitors in the global battle to develop platform-based PropTech solutions. Usually it is PropTech sectors from high-income countries, which are embedded in highly connected local innovation ecosystems that attract more venture capital funding and, hence, have a better chance to compete in their niche. At the same time, PropTech is not focused solely in large domestic markets, such as the U. S., China, India, Germany or the United Kingdom, a number of smaller markets, such as Switzerland, Australia or Ireland were also able to develop relatively large PropTech ecosystems.

5.3 | Data as key property technology (hypothesis 2)

At the core of PropTech is the technologies that are introduced by the start-ups in the sector. These technologies differentiate PropTech from the traditional real estate industry. As shown in the PropTech 2020 report, the most important technological trends in PropTech are construction technologies or ConTech, Smart Buildings, Real Estate FinTech, the Shared Economy, and Data Analytics. In this section we develop a data-driven categorisation of property technologies.

To do this, we utilise the technology categorisation system from Unissu to construct a network of technology combinations that are used by the PropTech firms. Figure 8 shows the total universe of PropTech firms (small nodes in the network) and the main technologies being used. The coloured links represent areas of the network that form densely

connected clusters of firm-technology combinations.

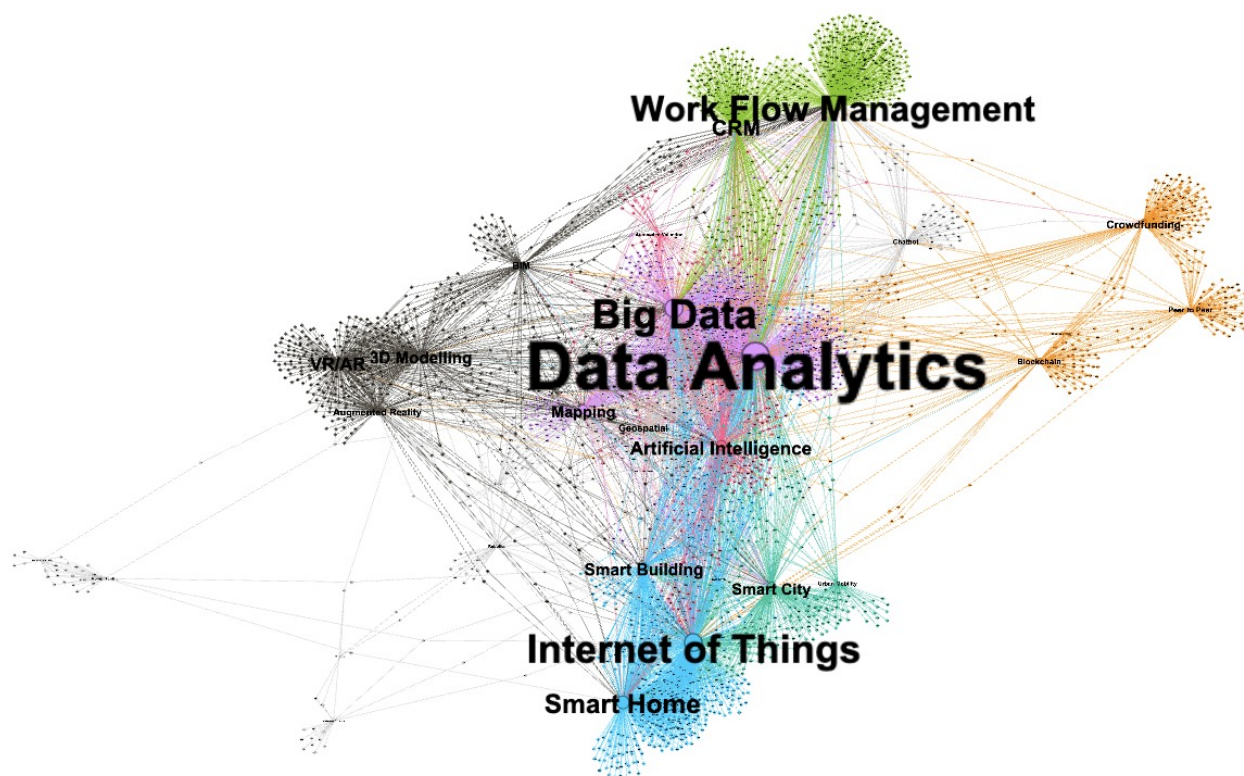


FIGURE 8 Property Technology bipartite network.

To better identify major technological clusters, we have applied association rule learning to the complete network dataset. The result is displayed in figure 9. It shows the most important 30 technologies used in PropTech. Each dot in the figure represents a technology, while the size of the dots corresponds to the total funding that went into the technology. Two technologies are connected if they tend to be used often together by firms. The association rule learning measure of lift defines the threshold for connections between two technologies as follows: all combinations with a lift value larger than 1 are displayed as arrows in the plot, identifying those tech combinations that are used together more often than by chance alone. The more often two technologies are used together, the darker is the line linking the technologies.

In order to identify clusters of technologies in the resulting property technology network, we apply the Louvain method for community detection. The technologies that are joined together in the resulting clusters represent those regions of the network that are more densely connected with each other than with other parts of the network.

The final network consists of six tech clusters, which largely correspond to the theoretical classification presented in the PropTech 2020 report. Data Analytics forms a distinct cluster at the core of the network, which is highly connected with the Smart Real Estate cluster that contains technology tags such as 'Smart Buildings' or 'Internet of Things'. Somehow connected to this core component of the network are the technologies around 3D Modelling and Virtual Reality. The tech cluster defining business processes — consisting of 'Work Flow Management', 'Chatbot', and 'CRM' tags — is only loosely connected to the other three clusters. Completely detached is the ConTech cluster, which includes the technologies 'Pre-Fabrication' and 'Modular Building'. Also detached is the Real Estate FinTech cluster,

Property Technology Network

Based on Unissu and CrunchBase data of 3,500 PropTech firms

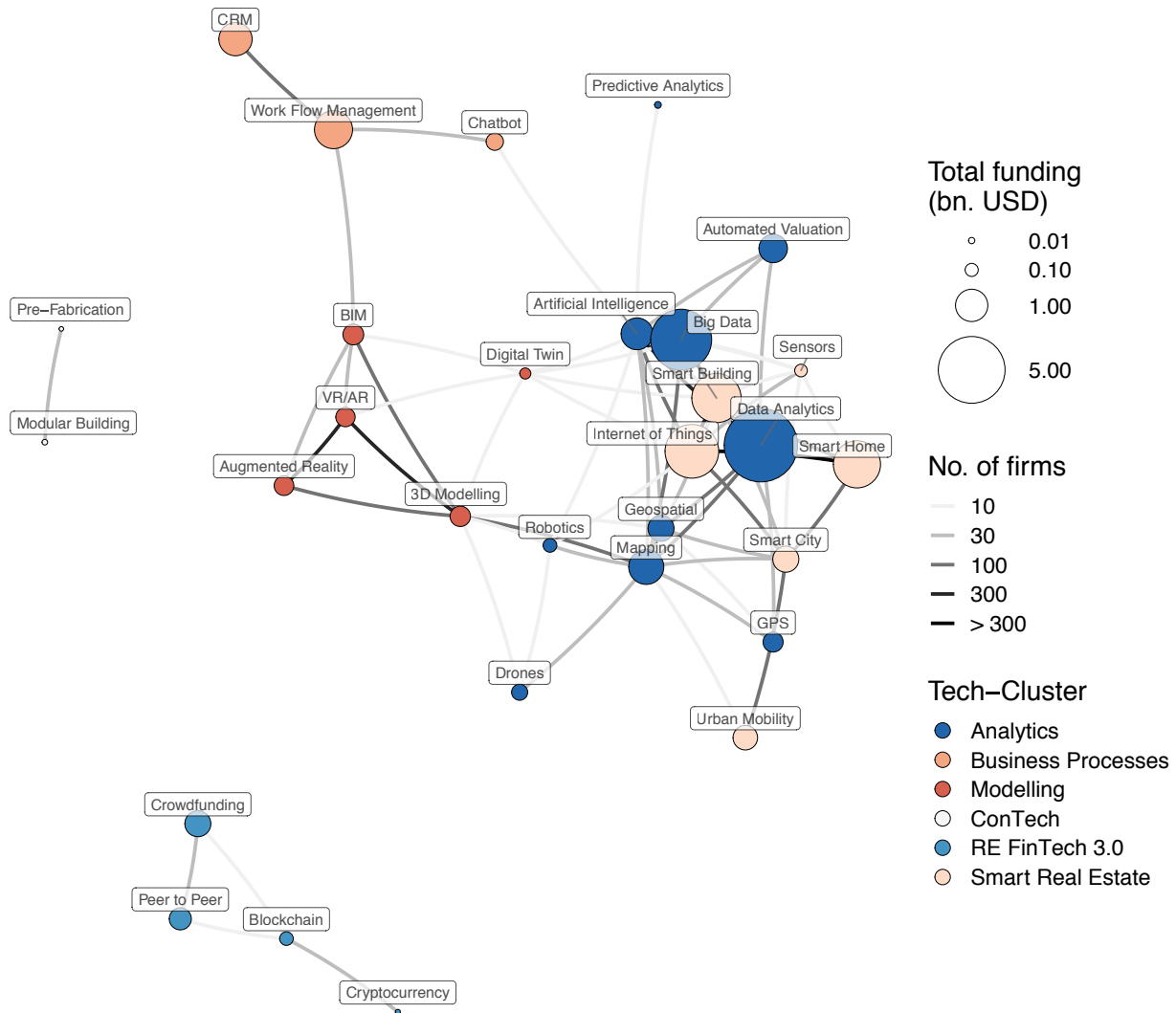


FIGURE 9 Property Technology network.

which contains the tags 'Crowdfuding', 'Peer to Peer', 'Blockchain', and 'Cryptocurrency'.

This categorisation does not imply that only the clustered or adjacent technologies can be combined together. Neither does it imply that the technologies placed at the periphery of the network have less potential to change the industry. However, the more central technologies are those that have already been successfully introduced into the real estate market. It is those central technologies that are at the core of PropTech: firms applying these technologies to real estate business problems are more likely to have attracted venture capital funding and they were more likely to bridge related technological domains.

The tech clustering that emerged from the data does not only represent mere statistical correlations, as the technologies clustered together are from similar tech domains. The two largest clusters, Smart Real Estate and Data Analytics deal with the collection of digital data and the value-generation from the data. The Modelling cluster includes those technologies that provide a digital representation of buildings, and the Real Estate FinTech cluster groups those tech-

nological solutions that are important for financing and transacting real estate.

To get a better understanding of the economic importance of the different technology groups, we turn towards the venture capital funding that each cluster obtained. Figure 10 shows the average funding per firm (in millions USD) on the vertical axis. Each dot represents a technology and its size corresponds to the number of firms in the dataset that use a technology. The total amounts of funding per cluster are shown in the inset. As indicated in the previous figure, it is the two largest clusters that also obtained highest funding per firm: 5.7 millions USD per Smart Real Estate firm, 6.2 millions USD per Data Analytics firm. The three clusters of Modelling, RE FinTech, and Business Processes got, on average, 2.1 to 2.8 million USD funding per firm and the ConTech sub-sector obtained just 300,000 USD funding on average.

There is therefore substantial variation in the market valuation of the different property technologies. Also, within each cluster there is high volatility. Overall, however, five individual technologies were able to attract the highest funding amounts. These were: 'Smart Building', 'Smart Home', and 'Internet of Things' in the Smart Real Estate cluster; and 'Data Analytics' and 'Big Data' in the Analytics cluster.

From this we can conclude that the most important property technologies in the current PropTech market are those that produce digital data and generate value from it. In comparison to these technologies, promising terms like Blockchain or Cryptocurrency or Modular Building remain peripheral. It is to be seen whether it will be possible to create sustainable business models that use these prominently discussed technologies. Based on the low number of firms using these technologies in combination with more accepted approaches, and due to the low funding raised, many promising technologies do not seem to be ready for commercial application in their current stage.

The grouping of property technologies in clusters can also be used to create a data-driven PropTech 'Hypecycle' Panetta (2018). Figure 11 displays the importance of each tech-cluster over time. To construct the yearly data points describing the hypecycle, the Property Technology network shown in figure 9 was created for each of the past twenty years, based on the founding year of the companies in the dataset. Then, we measure the Betweenness centrality (a measure of the importance of nodes in a network) of each technology in the tech clusters, calculate the cluster average and normalise by the yearly average betweenness centrality to make the networks comparable over time. The resulting measure shows the relative importance of each tech-cluster at any given year within the PropTech universe.

The resulting data-driven PropTech hypecycle shows distinct shifts in importance over the past two decades. The data analytics cluster, now the most important tech-group within PropTech, was very central already in the first stage of the observation period around 2000, before it steeply decreased in centrality between 2005 and 2008, and started to rise in importance from 2010 onwards. Business Processes, today a less central cluster, went through an opposite trend. Its centrality increased steadily up until 2007, when its importance started to decrease. In contrast, Modelling and Smart Real Estate showed less volatile dynamics, being relevant topics throughout the complete observation period. Real Estate FinTech and ConTech, on the other hand, were basically non-existent before 2010, but recently started to gain momentum; this is particularly the case for the Real Estate FinTech cluster.

From these considerations we can conclude that PropTech is essentially focused around a limited number of digital technologies that could successfully be applied to solve business problems in real estate. Big data, data analytics, and sensor technologies are the most promising technologies within the sector based on their utilisation and capitalisation. Other tech-clusters, such as those in the Real Estate FinTech group, have been trending recently, but it is yet to be seen whether they will become central to PropTech in the future. We can therefore conclude that data-related technologies are indeed at the core of PropTech.

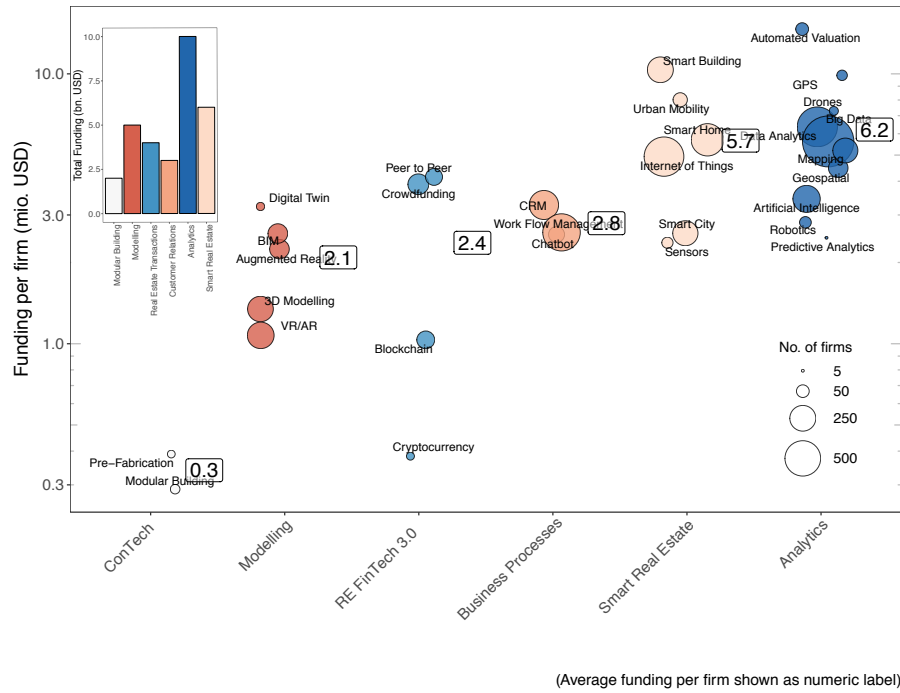


FIGURE 10 Size and funding per Property Technology cluster.

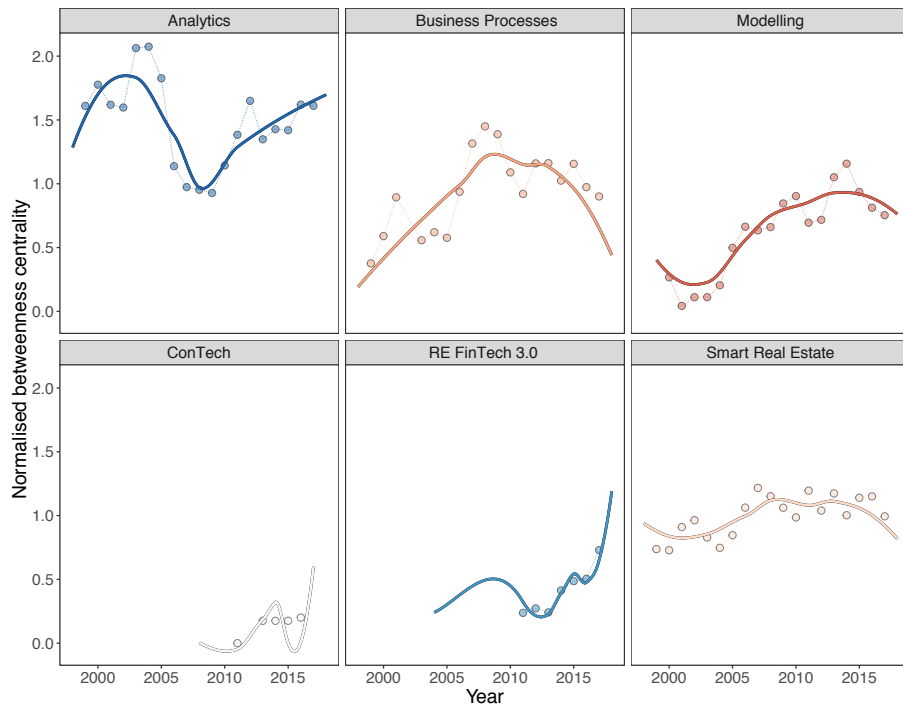


FIGURE 11 'Hypecycle' of Property Technologies.

5.4 | The winner takes all (hypothesis 3)

Now that we have derived the six core clusters of property technologies, we can use this categorisation system to investigate the third research hypothesis with regards to the competitive dynamics in PropTech. Based on the

TABLE 1 Acquisitions in different tech clusters.

	Tech-cluster	Acquisitions	Firms	Acquisitions / Firms
1	Analytics	554	1,369	0.40
2	Business processes	208	697	0.30
3	Smart Real Estate	226	837	0.27
4	Modelling	55	401	0.14
5	RE FinTech 3.0	17	268	0.06
6	ConTech	0	24	0.00

literature reviewed, it stands to reason that data-related business models and technologies should be characterised by the most intense competition for market share. It is in this sub-sector that the data-driven economics of scale are most relevant. In contrast, those technology-clusters that are more closely related to the material component of the real estate market are more likely to allow multiple competitors as the drive for natural monopolies and winner-takes-all dynamics are less pronounced in more manufacturing related industries such as ConTech.

Table 1 lists the number of acquisitions and firms per tech-cluster. Please note that, for simplicity, every firm was assigned to the cluster that contains most of the technologies a firm is applying. As hypothesised, it is indeed the data analytics cluster that has seen most acquisitions, not only in absolute terms (554 acquisitions), but also in relation to the number of firms in the cluster. In contrast, of the 401 firms in the more material cluster 'Modelling', there were only 55 acquisitions. None of the firms in ConTech has reported any acquisitions. From these observations we can conclude that the sub-sector of PropTech that most directly deals with the generation of digital data and its analysis aligns well with that part of the industry in which most acquisitions have occurred. This confirms our theoretical interpretation that the data-driven economies of scale provide an incentive to increase market share by buying competitors.

6 | CONCLUSION

6.1 | Summary

PropTech is currently challenging the real estate industry. It has been praised for introducing efficiency gains and innovation. While PropTech firms consider themselves as being digitally disruptive, they compare their role in the real estate industry with that of start-ups in FinTech, human resources or other sectors that went through a process of digitalisation. In such sectors, a number of characteristics of data-driven markets could be identified. In particular, these were the global scope of business models, the centrality of data-related businesses, and a fierce competition for market share.

In this paper, we investigate whether PropTech is turning real estate into a data-driven market. The quantitative findings from an analysis of more than 7,000 PropTech firms reveals that these trends are at work in PropTech. PropTech has become increasingly important in real estate. In recent years, there were more PropTech firms founded than non-PropTech real estate firms (according to the online platforms Unissu and Crunchbase). PropTech is clearly a global phenomenon, which is focused mostly in advanced economies in North America, Europe, China, and India.

Based on a network approach, we were able to identify a cluster of data analytics technologies that are at the core of the network of all property technologies. In comparison to other sectors, data analytics and related technologies were able to obtain substantially more venture capital funding, providing evidence for the hypothesis that data-driven

markets are characterised by a substantial need for external funding to finance the efforts to obtain larger market shares. This is emphasised by the observation that there are substantially more acquisitions in the data analytics sub-sector of PropTech than in any other part of the industry.

6.2 | Implications for the real estate industry

The findings presented in this study have important implications for users and owners of real estate. While the space-related materiality of the business of real estate might have protected the sector from digitalisation, the rise of PropTech shows that digital elements can substantially change the market logic of real estate.

The core business of real estate remains to be material or space-based, but the PropTech boom proves that essentially all information around using, managing, or owning real estate can be digitalised. At the core of PropTech is the datafication of real estate processes. For example, the search for a house has, thanks to websites such as Zillow, Rightmove or Zoopla, turned this search into a database query. Users apply a web-interface to search for a property that fits well to their preferences. In using such software, buyers of real estate provide valuable data about their market preferences. This logic applies to all kinds of information about real estate. Sensors and cameras allow managers to observe the utilisation of offices, while drones can be used to digitalise boundaries and to produce virtual representations of buildings.

All these innovations can potentially lead to substantial efficiency gains. However, at the same time, they change the whole fabric of the real estate market. Datafied markets are usually characterised by oligopolistic market structures, with a few firms or even monopolies offering the sole digital service available. The real estate market has not yet turned into a data monopoly, but the digital giants Amazon and Google have made first steps into the market. To prevent the accumulation of market power in a potentially datafied PropTech market and in order to benefit from efficiency gains associated with the introduction of digital technologies in the market, users and owners of real estate need to become aware of the value of the data they are generating in renting, buying, or managing real estate.

References

- Agrawal, R., Imieliński, T. and Swami, A. (1993). Mining association rules between sets of items in large databases. In: *Acm sigmod record*. Vol. 22. : ACM, pp. 207–216.
- Baum, A. (2017). *PropTech 3.0: the future of real estate*. Tech. rep., Saïd Business School, University of Oxford, Oxford, publisher: University of Oxford.
- Baum, A., Saull, A. and Braesemann, F. (2020). *PropTech 2020: the future of real estate*. Tech. rep., Saïd Business School, University of Oxford, Oxford, publisher: University of Oxford.
- Boeing, G. (2020). Online rental housing market representation and the digital reproduction of urban inequality. *Environment and Planning A: Economy and Space*, 52 (2), 449–468. Publisher: SAGE Publications Sage UK: London, England.
- Boeing, G., Besbris, M., Schachter, A. and Kuk, J. (2020). Housing Search in the Age of Big Data: Smarter Cities or the Same Old Blind Spots?.,
- Braesemann, F. (2019). The Evolution of Digital Technologies: A Network Perspective on Machine Learning. *Complex Network*, 10–12.
- Brynjolfsson, E. and Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, 358 (6370), 1530–1534. doi:10.1126/science.aap8062.

- Burrows, R. and Savage, M. (2014). After the crisis? Big Data and the methodological challenges of empirical sociology. *Big data & society*, 1 (1), 1–6. Publisher: SAGE Publications Sage UK: London, England.
- Ciuriak, D. (2018). The Economics of Data: Implications for the Data-driven economy,.Publisher: Chapter.
- De Mauro, A., Greco, M., Grimaldi, M. and Ritala, P. (2018). Human resources for Big Data professions: A systematic classification of job roles and required skill sets. *Information Processing & Management*, 54 (5), 807–817. doi:10.1016/j.ipm.2017.05.004.
- Easley, D. and Kleinberg, J. (2010). *Networks, crowds, and markets*. Vol. 8. : Cambridge university press Cambridge.
- Fields, D. (2019). Automated landlord: Digital technologies and post-crisis financial accumulation. *Environment and Planning A: Economy and Space*, 1–22.Publisher: SAGE Publications Sage UK: London, England.
- Fields, D. (2019). The Politics of Digital Transformations of Housing. *Planning Theory & Practice*, 20 (4), 578–582. Publisher: Taylor & Francis.
- Forman, C., Goldfarb, A. and Greenstein, S. (2018). How Geography Shapes-and is Shaped by-the Internet. *The New Oxford Handbook of Economic Geography*, 269. Publisher: Oxford University Press Oxford, UK.
- Gordon, E. (2007). Mapping digital networks from cyberspace to Google. *Information, Communication & Society*, 10 (6), 885–901. Publisher: Taylor & Francis.
- Keen, A. (2015). *The Internet is not the answer*. : Open Road+ Grove/Atlantic.
- Kenney, M. and Zysman, J. (2016). The rise of the platform economy. *Issues in science and technology*, 32 (3), 61. Publisher: Issues in Science and Technology.
- Landau-Ward, A. and Porter, L. (2019). Digital Innovations, PropTech and Housing–the View from Melbourne. *Planning Theory & Practice*, 20 (4), 582–590. Publisher: Taylor & Francis.
- Lee, M. (2010). A political economic critique of Google Maps and Google Earth. *Information, Communication & Society*, 13 (6), 909–928. Publisher: Taylor & Francis.
- Liggio, B. and Haggerty, J. (2019). Mergers and acquisitions: Impact on the real estate technology revolution. *Corporate Real Estate Journal*, 9 (2), 171–184. Publisher: Henry Stewart Publications.
- Loukissas, Y. A. (2017). Taking Big Data apart: local readings of composite media collections. *Information, Communication & Society*, 20 (5), 651–664. Publisher: Taylor & Francis.
- Manning, S., Larsen, M. M. and Kannothea, C. G. (2017). Global sourcing of business processes: History, effects, and future trends. *The new Oxford handbook of economic geography*, 407–435.Publisher: Oxford University Press Oxford.
- Mayer-Schönberger, V. and Ramge, T. (2018). *Reinventing capitalism in the age of big data*. : Basic Books.
- Meyer-Schonberger, V. and Cukier, K. (2012). *Big Data: A Revolution that will Transform How we Work, Think and Live*. : New York: Mariner.
- Panetta, K. (2018). *5 Trends Emerge in the Gartner Hype Cycle for Emerging Technologies, 2018*.
URL <https://blogs.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018/>
- Porter, L. (2019). Planning, Land and Housing in the Digital Data Revolution. *Planning Theory & Practice*, 20 (4), 575–578. Publisher: Taylor & Francis.
- Rogers, D. (2016). Uploading real estate: Home as a digital, global commodity. In: *Housing and Home Unbound*, pp. 37–52.

- Rogers, D. (2017). Becoming a super-rich foreign real estate investor: Globalising real estate data, publications and events. In: *Cities and the Super-Rich*, pp. 85–104.
- Rogers, D. and Koh, S. Y. (2017). *The globalisation of real estate: The politics and practice of foreign real estate investment*. : Taylor & Francis.
- Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, 6 (1), 2053951718820549. Publisher: SAGE Publications Sage UK: London, England.
- Sassen, S. (2012). Interactions of the Technical and the Social: Digital Formations of the Powerful and the Powerless. *Information, Communication & Society*, 15 (4), 455–478. Publisher: Taylor & Francis.
- Saull, A., Baum, A. and Braesemann, F. (2020). Can Digital Technologies Speed up Real Estate Transactions? *Journal of Property Investment & Finance*, Forthcoming.
- Shaw, J. (2018). Platform Real Estate: theory and practice of new urban real estate markets. *Urban Geography*, 1–28. Publisher: Taylor & Francis.
- Solomont, E. B. (2020). *Is it over for iBuying? Pandemic causes major players to hit pause*.
URL <https://therealdeal.com/miami/2020/04/13/is-it-over-for-ibuying-pandemic-causes-major-players-to-hit-pause/>
- Sundararajan, A. (2016). *The sharing economy: The end of employment and the rise of crowd-based capitalism*. : Mit Press.
- Yeung, H. W.-C. (2018). The Logic of Production Networks. *The New Oxford Handbook of Economic Geography*,.Publisher: Oxford University Press Oxford, UK.