

When activities connect: Sequencing, network analysis, and energy demand modelling in the United Kingdom

Eoghan McKenna^{A,1}, Sarah Higginson^A, Tom Hargreaves^B, Jason Chilvers^B and Murray
Thomson^A

^A Centre for Renewable Energy Systems Technology (CREST),
Wolfson School of Mechanical, Manufacturing and Electrical Engineering
Loughborough University, LE11 3TU, UK

^B Science, Society and Sustainability (3S) Research Group, School of Environmental Sciences,
University of East Anglia, Norwich, NR4 7TJ, UK

¹ Present address: UCL Energy Institute, University College London (UCL), Central House, 14
Upper Woburn Place, London, WC1H 0NN. e.mckenna@ucl.ac.uk

Abstract

This work applies a network analysis technique to the study of real and synthetic residential activity data commonly used in activity and energy demand research.

UK Time Use Survey activity diaries are converted into network graphs of activity sequences. Differences between weekday and weekend networks are compared using network metrics: size, density, centrality and homophily. The results show that the weekday activity sequence network is smaller, less dense, more central and has lesser homophily than the weekend network.

The technique is applied to test the validation of a model of residential active occupancy in buildings that uses a first-order Markov chain technique to generate synthetic data. The results show that the synthetic data reproduces *relative* differences between the network metrics for weekdays and weekends but the differences between real and synthetic data are statistically significant and greater or comparable to the differences observed between real weekday and weekend data. The first-order Markov chain technique fails to capture important characteristics of the sequence network that are present in the real data.

The analysis technique presented here can be used to improve the testing and validation of such models in future, as well the comparative analysis of sets of aggregated activity data for periods of known difference in energy demand.

Keywords: energy demand; time use diary; activity; network theory; network analysis; Markov chain; bottom-up modelling.

1. Introduction

To secure a low-carbon future, energy consumption behaviour will need to change. In particular, there is value in securing energy demand that is flexible in time (flexible demand), to reduce peak demand and help integrate intermittent sources of renewable energy into power systems [1]. In assessing the potential of flexible demand it is essential to understand the activities that drive energy demand in terms of their timing, energy intensity and possible flexibility [2,3].

The study of activities and their relationship with energy demand is an active area of research with numerous sub-fields. The present work aligns and contributes to several of these: the comparative analysis of sets of activity data [2–4]; the analysis of sequences of activities [5,6]; and ‘bottom-up’ models of energy demand that are based on activity data [7–11].

In this work we demonstrate a technique developed by Cornwell [12] that: converts activity data into network graphs in a way that retains information about the sequences of activities; calculates metrics for these networks that contain information about the characteristics of the aggregated sequences such that they are meaningful global descriptions of the underlying data; and compares differences in these metrics between sets of activity data that are the result of differences in the structure of the sequences present in the underlying data. With two examples, we show that this technique can add value to energy demand research by revealing novel insights into data and modelling techniques that are used in the sub-fields mentioned above.

In the first case, we compare residential activity data between weekdays and weekends, using data from the UK Time Use Survey [13]. Energy demand profiles for weekdays and weekends are known to be significantly different from each other, and previous research on time use survey data has shown that this difference is reflected in the nature of activities performed on these days e.g. what activities occur, and their timing and synchronisation [2]. We add insight to this literature by revealing differences between the two types of day in terms of the characteristics of the aggregated *sequences* of activities that underpin them.

In the second case, we compare real activity data with synthetic activity data generated by an algorithm that is commonly used in the field of energy demand modelling (the first-order Markov chain technique). Previous research has validated this technique by demonstrating the synthetic data generated by the technique was sufficiently comparable to real data for the purposes of this class of energy demand model [8,14]. Contrary to previous research, our results show that the synthetic data differs considerably from the real data in ways that suggest the technique fails to capture characteristics of the activity sequences present in the real data. This work adds value to the field of bottom-up activity-based energy demand modelling by providing novel constructive criticism of a commonly used modelling algorithm and demonstrating the use of a new technique that can be used in future for the validation of the output of such models.

2. Literature review

Activity data has been the subject of analysis within energy demand research community before. This section provides a brief summary of the relevant literature.

2.1. Analysis of activity data within energy demand research

Torrìti et al. [3] analysed 10-min activity diary data, questionnaires and 1-min GPS data for 175 days and 153 participants to better understand the synchronisation of activities, the greenhouse gas intensity of activities and the potential flexibility to shift activities throughout the day. They found greenhouse gas intensity and flexibility to shift activities varies throughout the day, with morning peaks characterised by high levels of synchronisation, and evening peaks characterised by low levels of synchronisation and high spatial mobility.

Anderson and Torrìti [2] analysed and compared UK time-use survey data from 1974 to 2014 to identify changes to the timing of aggregated activities, with a focus on changes in energy-intensive activities and to patterns of activities during the evening peak period (4pm to 8pm). Their findings suggest that the changes in social structures such as labour market participation and their knock-on impact on patterns of activities may have contributed to the observed changes in electricity demand over this period.

Sekar et al. [15] analysed and compared the American Time Use Survey between 2003 and 2012 to identify changes in the timing that high-level aggregated activities are performed and to estimate the impact of these changes on aggregated energy demand. Their findings indicate that more time at home and less time spent travelling and at work reduced national energy demand by 1.8%.

McKenna et al. [6] used network theory techniques to analyse 2000-2001 UK Time Use Survey. They used network metrics of degree and centrality to identify that personal care, household and family care, and travel are categories of activities are the most connected and suggested these act as 'hubs' or 'anchors' – particularly central or important activities – within the overall network of activities.

Lorincz et al. [5] analysed the timing and sequences of activities in the 2014-2014 UK Time Use Survey. They used network analysis techniques to create network graphs from individual activity diaries to identify the most frequent sequences of activities during periods of peak national energy demand. Using the degree network metric, they identified that travel activities have the highest number of connections, followed by food preparation, eating and mass media activities.

A key limitation of these studies are the lack of combined observations of activities and energy demand. Such studies are limited by the assumptions they must make regarding the link between activities and energy consumption. For example, what appliances are being used when an activity is performed, and how the energy-intensity of activities may vary between or within households. Grunewald and Diakonova [4] bridged these gaps by developing new methods to collect and analyse household activity and electricity consumption data. Based on data for 293 households and over 16,000 reported activities combined with 1-min concurrent electricity readings, they identified typical electricity intensities of household activities. They developed the concept of marginal electricity demand which allows the contribution of specific activities to electricity demand to be isolated. Their results show, for example, that hot meal related activities have strong contributions to electricity demand, while cold meals have the opposite effect.

This paper contributes to this field of research by demonstrating the value of a technique which explicitly captures properties of the sequences of activities, facilitating the comparison of these characteristics between sets of activity data, in particular where these are observed to have associated significant differences in energy demand profiles.

2.2. Bottom-up activity-based energy demand modelling

Time use activity data is also used in the field of energy demand modelling. It is used in 'bottom-up' activity-based models [10,11,14,16] which are used to simulate energy demand in residential buildings at high resolution, and which are typically used in low-voltage network analyses [17] and urban energy systems modelling [18].

A defining feature of such models is that they use real activity data, typically large national surveys such as the UK Time Use Survey, to calibrate an algorithm which is then used to create synthetic activity data at scale. The use of synthetic data rather than real data within this class of model stems, in part, from the historic lack of good quality real data at the time of the development of early models, and the need therefore to generate large amounts of 'realistic looking' data given limited availability of real data. There was also a general desire for such models to be 'self-contained' and open-source, and not require the accompaniment of large datasets with their distribution, which may have restrictive licencing conditions regarding onward distribution. Finally, there is a value in models having the functionality of producing realistic synthetic data at scale to suit the needs of whatever scenario is of interest to the modeller.

A key requirement, however, is that the output of such models (the synthetic data) provides an accurate enough representation of the real data which the algorithm was calibrated on [8]. As a result, these models are generally 'validated' by comparing the synthetic data they produce with the original real data on which the model was based.

By way of example, the 'first-order Markov chain' algorithm, which is the subject of scrutiny in this paper, was first validated in terms of state probabilities [14,19,20]. This means that, given a sufficiently large number of runs of the model, the probability that an activity occurs throughout the day is the same in the model as it is in the real data. Subsequent research

then criticised the algorithm's ability to accurately capture state (activity) durations [11]. This refers to whether, given a sufficiently large number of runs, the probability distribution of the duration of activities matches between the model and real data. Partly in response to this criticism, McKenna et al. [8] quantified the discrepancy in activity durations of the model and found that "the first-order Markov chain technique produces output that matches closely the original data, with good agreement in terms of state probabilities and state durations".

In this work we contribute to this literature by taking the same data used by McKenna et al [8] to validate the first-order Markov chain model in terms of state probability and durations, and analyse it using a network analysis technique to test whether the algorithm can be validated in terms of the characteristics of the sequences of activity data it produces.

3. Method

3.1. Sequence network analysis of residential activity diaries

Network theory is a broad field of research devoted to the scientific analysis of the structure and dynamics of networks, including technical, information, biological and social networks [21–23]. It has proven to be a useful method for understanding networks that consist of flows, traffic, or sequences, such as the diffusion of viruses or innovations through social networks [24], or the transcriptional regulation networks that regulate gene expression in cells [25].

In this work, network theory techniques are applied to sequence data to produce and analyse sequence networks [26]. Sequence networks are networks comprising multiple overlapping individual chains of sequences. We adopt the method developed by Cornwell [12] of converting sets of individual activity chains into sequence network graphs and analysing the resulting networks using network analysis metrics. The method is described in detail in [12] and briefly summarised here.

Figure 1 shows a simple illustrative example of how multiple sequences of activity data are converted into a network graph for subsequent analysis. Activity sequence data is structured as a simple two-dimensional array. Each row represents a separate sequence, with columns representing consecutive time periods. The data analysed in this paper consists of 10-minute resolution activity data for a 24-hour period, so each sequence consists of 144 elements. Each element contains a code representing the activity that individual performed in that 10-minute period.

1. A simple example of two sequences of activities.

	Time 1	Time 2	Time 3	Time 4
Sequence 1	Activity A	Activity A	Activity A	Activity C
Sequence 2	Activity B	Activity A	Activity A	Activity B

2. Converting the sequence data into an 'adjacency list'

	Source node	End node	Weight
Edge 1	1_A	2_A	1
Edge 2	2_A	3_A	2
Edge 3	3_A	4_C	1
Edge 4	1_B	2_A	1
Edge 5	3_A	4_B	1

3. Converting the adjacency list into a directed network graph

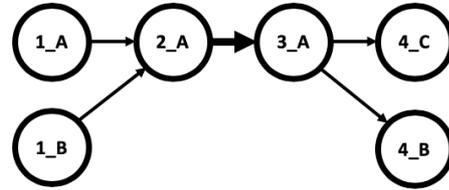


Figure 1 – illustrative example of steps involved in converting activity sequence data into a network graph.

Network graphs consist of nodes and edges which are the connections between the nodes. The nodes of a network of activity sequences consist of an activity and the time at which it occurred. In the simple example there are six distinct activity-time combinations and therefore six nodes in the graph. The two sequences ‘overlap’ during times 2 and 3 as the same activity occurred at the same times for both.

The edges, or connections, of the graph represent the observed sequences of activities. Node 1_A is connected to 2_A because in Sequence 1 Activity A occurred at Time 1 and was followed by Activity A at Time 2. The sequence data is converted in this way into an ‘adjacency list’ which is a list of edges. The adjacency list includes information about the ‘weight’ of each edge. In this case, both Sequence 1 and Sequence 2 contain the same transition between time 2 and 3. This connection is therefore given a weight of 2. In this case, therefore, higher weights represent paths that are more common. Because the data represents sequences, there is a meaningful direction to the connections between nodes. The edges are directed, and these therefore produce a *directed graph*.

The adjacency list can then be converted into a directed network graph and analysed using numerous publicly available network software. Here the NetworkX package [27] is used with Python.

3.2. Data

The data for this paper’s first case-study is the 2000-2001 UK Time Use Survey [13]. This consists of 20,981 diary entries of 24-hour diaries of people’s reported activities, at 10-minute resolution. Each person who participated completed a weekend and a weekday diary entry. In total, there are 253 activities logged under ten broad categories, such as ‘personal care’, or ‘mass media’. The weighted sample is intended to be nationally representative. Weightings are not used in this work as the focus is on the comparative difference between weekday and weekend activities. The UK Time Use Survey reports primary and secondary activities. Only the former are used here.

The data was processed to ensure that an equal number of weekend and weekday diaries and that diaries were only included where a participant completed both. This resulted in a final set of 10,338 diaries for each set.

The data for the second case-study was used previously to validate the output of a model that produces synthetic data of residential ‘active occupancy’ in buildings [8]. Active occupancy is defined as when an occupant is at home and not asleep. The model generates time-series active occupancy data at 10-minute resolution for a 24-hour period for a specified number of occupants in a specified number of residential buildings. The data is different from the individual activity diaries of the UK Time Use Survey in two ways (see Figure 2). First it is a more basic representative of activity, in that it only indicates whether occupants are at home and not asleep. This is represented by two digits, the first for the occupants’ presence in the building and the second for their activity. “10” means one occupant at home and asleep, while “11” means one occupant at home and awake (active). Second, it is aggregated across all members of a household. For example, for a three-person household, the UK Time Use Survey data would consist of three activity diaries. By contrast the model data consists of one (aggregated) active occupancy sequence.

1. Example of active occupancy data for a three-person household.

	Time 1	Time 2	Time 3	Time 4
Occupant 1	10	10	11	01
Occupant 2	10	10	10	11
Occupant 3	10	10	10	10
Aggregated	30	30	31	22

2. Converting the aggregated sequence data into an ‘adjacency list’

	Source node	End node	Weight
Edge 1	1_30	2_30	1
Edge 2	2_30	3_31	1
Edge 3	3_31	4_22	1

3. Converting the adjacency list into a directed network graph

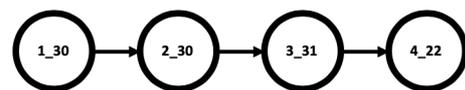


Figure 2 – example of aggregated active occupancy data and its conversion to a network graph.

The UK Time Use Survey is converted into aggregated activity occupancy sequences and split into weekend data and weekday data. A random selection of 5000 sequences from each are chosen and compared to 5000 sequences of weekend and weekday synthetic data which has been generated to have the same distribution of occupants as in the real data.

Aggregated sequence data is then converted into adjacency lists and then directed network graphs for analysis.

3.3. Metrics

This section describes the metrics used in this paper to analyse and compare the network graphs created using the data described above. The metrics are *network size*, *density*, *centrality* and *homophily*. These are common metrics in network analysis. Cornwell [12] describes these metrics and their relevance to activity sequence networks, and which we summarise below.

Network size refers to the number of unique nodes in a network. For activity sequence networks the size describes the number of unique observations of activities and the time they occur at. A larger network is one where more activities were observed and gives an indication of the diversity of activities that occurred.

Network density is the fraction of the total maximum number of possible connections that are actually observed in the network. There are 253 unique activities reported in the UK Time Use Survey, so the maximum possible number of connections is $253 \times 253 \times 143 = 9,153,287^1$. Network density is calculated by counting the number of unique edges observed in the network and dividing it by this maximum. It is a measure of the diversity of sequences present in the network. A denser network is one where there are more alternative pathways between time periods by different people.

Centrality refers to how dominant certain pathways or nodes are within a network. Central activity nodes are those that are highly connected to others. Central connections or edges are those that appear in many individual sequences i.e. those with high 'weight'. There are multiple ways of calculating centrality, here we use edge weight as a measure of (path) centrality and compute the mean and standard deviation of the distribution of edge weights for each network.

Homophily refers to the likelihood that a node is connected to another node of the same type. Here this refers to when an activity in one time period is connected to the same activity in an adjacent time period. It is a measure of the extent to which activities tend to endure in unbroken lasting sequences. A simple measure of homophily used here is the proportion of times an activity in one time period is followed by the same activity in the following time period. This is calculated for each sequence and the mean of the distribution is taken as the measure of the homophily of the network.

¹ Each time period can have a maximum of 253 activity nodes and each of these can be connected to a maximum of 253 activity nodes for the next time period. There are 144 10 minute time periods in a day and so there are 143 of such possible connections between activity nodes in consecutive time periods.

4. Results

4.1. Weekday activities versus weekend activities

Table 1 shows the results of calculating the network metrics described previously for the UK Time Use Survey and compares these for weekdays and weekends. The comparison of network size shows that the weekend network is larger, meaning there is a greater diversity of activities performed on these days. The weekend network is also denser, meaning there are a greater number of alternative pathways between time periods by different people on weekends than on weekdays. When changes of activity do happen, there is a greater variety of change of activity on the weekend. In line with this, the centrality metric shows that the weekend network is less central than the weekday as it has smaller edge weights. This means that on weekdays more people tend to be following the same sequences of activities than on the weekend. The homophily measure is greater for the weekend network meaning that while weekends have greater diversity of activities (network size) and changes to activities (density) activities occurring on weekends tend to endure longer than those that occur on weekdays.

Table 1 – a comparison of network analysis metrics for weekday and weekend UK Time Use Survey activity sequence data.

	Weekday	Weekend
Number of diaries	10338	10338
Network size	20235	21220
Network density (%)	1.13	1.20
Centrality: mean edge weight (and std. dev.)	14.2 (217.5)	13.5 (209.9)
Homophily (%)	84.00	84.09

4.2. Real versus synthetic active occupancy data

Table 2 shows the results of calculating the network metrics described previously for real and synthetic active occupancy networks broken down by weekday and weekend. Looking first at the real data, the metrics are different than those for the activity data shown in Table 1. This is to be expected as the data is different (see 3.2). Nonetheless the comparative differences between weekday and weekend networks are the same as for the activity data: the weekend network is larger, denser, less central, and has greater homophily. This indicates that while active occupancy is different from activity data it retains the comparative difference between weekdays and weekends from a sequence network perspective.

Table 2 – network analysis metrics for real and synthetic active occupancy sequence networks for weekday and weekend.

	Weekday (real data)	Weekend (real data)	Weekday (synthetic data)	Weekend (synthetic data)
Number of sequences	5000	5000	5000	5000
Network size	5553	5723	4554	4846
Network density (%)	8.77	9.25	7.03	7.48
Centrality: mean edge weight (and std. dev.)	44.0 (172.4)	41.7 (166.4)	54.9 (192.3)	51.6 (184.8)
Homophily (%)	91.41	91.76	91.47	91.84

The synthetic data also captures the expected differences between weekday and weekend networks: the synthetic weekend network is larger, denser, less central, and has greater homophily than the synthetic weekday network.

There are however considerable differences between the metrics for the real and synthetic networks. The synthetic networks are smaller, less dense, and more central than the real networks. The synthetic data does not capture the full diversity of nodes (active occupancy states) or connections (state transitions) and synthetic sequences are more likely to follow central pathways. By contrast, while the synthetic networks have greater homophily than their real counterparts, the difference is small and, arguably, this metric is reasonably well reproduced by the synthetic data.

4.3. Test of statistical significance

We test the statistical significance of these differences using a subsampling ‘bootstrap’ technique. Subsamples from each sample (weekday real data, weekend real data, weekday synthetic data, weekend synthetic data) are chosen each comprising 1000 sequences randomly selected without replacement. Network graphs are created using the subsamples and the network metrics are calculated. This is repeated 1000 times, resulting in distributions for each metric for each sample. Of these, the mean and 99% confidence interval are calculated and an unequal independent sample t-test is calculated to test for statistically significant difference between sample means (Table 3). For real data, the weekend sample is tested against weekday, and the synthetic samples are tested against their equivalent day real samples.

Table 3 – results for 1000 sub-samples of 1000 sequences randomly selected without replacement. Results show mean values with 99% confidence interval provided in brackets. Asterisks indicate statistically significant difference (** p-value < 0.01, * p-value < 0.05, * p-value < 0.1) between mean values for weekday and weekend for real data, and between real and synthetic data for days of the same type.

	Weekday (real data)	Weekend (real data)	Weekday (synthetic data)	Weekend (synthetic data)
--	--------------------------------	--------------------------------	-------------------------------------	-------------------------------------

Network size	4045 (4041, 4050)	4226*** (4222, 4231)	3850*** (3843, 3857)	4059*** (4051, 4067)
Network density (%)	4.65 (4.65, 4.66)	4.87*** (4.87, 4.87)	4.41*** (4.40, 4.42)	4.62*** (4.61, 4.62)
Centrality: mean edge weight	16.6 (16.6, 16.6)	15.8*** (15.8, 15.8)	17.5*** (17.5, 17.5)	16.7*** (16.7, 16.8)
Homophily (%)	91.40 (91.40, 91.41)	91.76*** (91.76, 91.76)	91.47*** (91.46, 91.48)	91.83*** (91.82, 91.84)

The results show that some metrics are affected by the number of sequences used to form the network graph. Network size, density and centrality (mean edge weight) are all smaller for graphs made from fewer sequences (in this case 1000 rather than 5000) and this makes sense as these are metrics that should scale with the number of sequences. By contrast, homophily values are replicated, which is again to be expected.

The tests of statistical significance show that the relative differences between graphs that were indicated by the results in the previous section are reproduced here at the 99% confidence level. With the exception of homophily, the differences between synthetic and real data are comparable or greater than the differences between real weekday and weekend data. The results show that while homophily is similar between synthetic and real networks, the differences are in fact statistically significant.

5. Discussion

5.1. Activity sequences and their impact on energy demand

The results show that the (real) weekday activity sequence network is smaller, less dense, more central and has lesser homophily than the (real) weekend network. Weekdays are when most people work and there is a greater level of synchronisation in people's activity sequences. Most people do not work on weekends and can pursue other, more varied, activities and can afford to spend more time doing these.

Energy demand profiles for weekdays and weekends are known to be significantly different from each other, and previous research on time use survey data has shown that this difference is reflected in the nature of activities performed on these days e.g. what activities occur, and their timing and synchronisation [2].

We add insight to this literature by revealing differences between the two types of day in terms of the characteristics of the aggregated *sequences* of activities that underpin them. We do not claim these differences *cause* differences in energy demand between weekday and weekend. However if we accept the proposition that the sequencing of aggregated activities has an effect on the resulting patterns of aggregated energy demand, then we have demonstrated a technique and metrics that can quantify differences between groups of aggregated activity sequences, and we have shown that there are considerable differences in such metrics between types of day that are known to have considerable differences in energy demand (weekday vs weekend). Knowledge about the association between these metrics and energy demand profiles can be used in future research to create and test hypotheses about the relationship between activity sequences and energy demand e.g. are activity sequence networks that are less central more flexible to change (a question which is relevant to flexible demand studies)?

5.2. Energy demand model validation

The results show that the synthetic (aggregated active occupancy) data produced by an established modelling algorithm (first-order Markov chain) manages to reproduce the *relative* differences between weekdays and weekends but performs less well in absolute terms. With the exception of the homophily measure, the differences between real and synthetic data are comparable or greater than the differences between weekdays and weekends. From a sequence network analysis perspective, the algorithm fails to capture important characteristics that are present in the real data.

The homophily is similar between synthetic data and real data. Homophily is a measure that is related to the duration of states, and the algorithm has previously been validated in terms of state durations [8], so these results confirm this characteristic of the data is relatively well-represented. The differences between homophily for synthetic and real data are nonetheless statistically significant, indicating that algorithm does not capture these characteristics perfectly.

The algorithm produces synthetic active occupancy sequence networks that are smaller, less dense, and more uniform and synchronised than real active occupancy sequence networks. It has been recognised that the first-order Markov chain algorithm '*under-represents the beginning and end of the range of state durations and over-represents the middle.*' [8] This is because the algorithm is based on transition probabilities that are calculated on means of the input data and that the distribution of states in the output data are determined by the random outcome of the Markov process. While this has been shown to have limited impact on state probabilities, and a relatively minor impact on state durations [8], our results show that it has a more considerable impact on the validity of the synthetic data when analysed from a sequence network analysis perspective.

Whether this error in the validity of the synthetic data matters for the purposes of the class of model that uses this data as an input for simulating energy demand is unknown and out of scope of this paper. Our results do suggest however that developers of such models should be aware that there are additional limitations of this type of algorithm, and this in turn may encourage them to improve or change their methods if current approaches fail to capture the complexity of the real data. Finally, this work demonstrates the use of a new technique and metrics which model developers can use in future for the testing and validation of such models.

6. Conclusion

Residential energy demand can be viewed as a by-product of people performing activities. To inform efforts to reduce energy demand or make it more flexible, the study of activities and their relationship with energy demand is an active area of research. This work applies a network analysis technique to the study of data and a modelling technique that are widely used in the field of activity and energy demand research.

Activity diaries taken from the UK Time Use Survey data are converted into activity sequence network graphs. The differences between weekday and weekend networks are compared using common network metrics: size, density, centrality and homophily. The results show that the weekday activity sequence network is smaller, less dense, more central and has lesser homophily than the weekend network. The results are intuitive. Weekdays are when most people work and there is a greater level of synchronisation in people's activity sequences. Most people do not work on weekends and can pursue other, more varied, activities and can afford to spend more time doing these.

We add insight to the literature on comparative analysis of aggregated activity data for periods of known difference in energy demand by revealing differences between two types of day with different demand profiles in terms of the characteristics of the aggregated *sequences* of activities that underpin them.

The same network analysis technique is applied to test the validation of a model of residential active occupancy in buildings. The model uses a first-order Markov chain technique to generate synthetic data which has previously been validated in terms of state probabilities and state durations by comparing the synthetic data with real data.

The results show that the synthetic data produced by the model manages to reproduce *relative* differences between the network metrics for weekdays and weekends but performs less well in absolute terms. With the exception of the homophily measure, the differences between real and synthetic data are statistically significant and comparable or greater than the differences observed between real weekday and weekend data. From a sequence network analysis perspective, the first-order Markov chain technique fails to capture important characteristics that are present in the real data. The analysis technique presented here can be used to improve the testing and validation of such models in future.

7. Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council, UK, within the Realising Transition Pathways project (EP/K005316/1).

8. References

- [1] N. Eyre, S.J. Darby, P. Grünewald, E. McKenna, R. Ford, Reaching a 1.5°C target: Socio-technical challenges for a rapid transition to low-carbon electricity systems, *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* 376 (2018). <https://doi.org/10.1098/rsta.2016.0462>.
- [2] B. Anderson, J. Torriti, Explaining shifts in UK electricity demand using time use data from 1974 to 2014, *Energy Policy*. 123 (2018) 544–557. <https://doi.org/10.1016/j.enpol.2018.09.025>.
- [3] J. Torriti, R. Hanna, B. Anderson, G. Yeboah, A. Druckman, Peak residential electricity demand and social practices: Deriving flexibility and greenhouse gas intensities from time use and locational data, *Indoor Built Environ.* 24 (2015) 891–912. <https://doi.org/10.1177/1420326X15600776>.
- [4] P. Grünewald, M. Diakonova, The specific contributions of activities to household electricity demand, *Energy Build.* 204 (2019). <https://doi.org/10.1016/j.enbuild.2019.109498>.
- [5] M.J. Lorincz, T. Yunusov, J. Torriti, Exploratory analysis of family-related activities during peak electricity periods, (2019).
- [6] E. McKenna, S. Higginson, T. Hargreaves, J. Chilvers, M. Thomson, Exploratory analysis of time-use activity data using network theory, in: Lancaster University, 2016. <https://www.researchgate.net/publication/304310048> (accessed January 24, 2020).
- [7] J.L. Ramírez-Mendiola, P. Grünewald, N. Eyre, Residential activity pattern modelling through stochastic chains of variable memory length, *Appl. Energy*. 237 (2019) 417–430. <https://doi.org/10.1016/j.apenergy.2019.01.019>.
- [8] E. McKenna, M. Krawczynski, M. Thomson, Four-state domestic building occupancy model for energy demand simulations, *Energy Build.* 96 (2015). <https://doi.org/10.1016/j.enbuild.2015.03.013>.
- [9] I. Richardson, M. Thomson, D. Infield, C. Clifford, Domestic electricity use: A high-resolution energy demand model, n.d. <https://dspace.lboro.ac.uk/> (accessed January 24, 2020).
- [10] D. Aerts, J. Minnen, I. Glorieux, ... I.W.-B. and, undefined 2014, A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison, Elsevier. (n.d.). <https://www.sciencedirect.com/science/article/pii/S0360132314000304> (accessed January 24, 2020).
- [11] U. Wilke, F. Haldi, J. Scartezzini, D.R.-B. and Environment, undefined 2013, A bottom-up stochastic model to predict building occupants' time-dependent activities, Elsevier. (n.d.). <https://www.sciencedirect.com/science/article/pii/S0360132312002867> (accessed January 24, 2020).

- [12] B. Cornwell, Network Analysis of Sequence Structures, in: 2018: pp. 103–120. https://doi.org/10.1007/978-3-319-95420-2_7.
- [13] Office for National Statistics, Ipsos-RSL, United Kingdom Time Use Survey, 2000, UK Data Serv. (2003). <https://doi.org/10.5255/UKDA-SN-4504-1>.
- [14] I. Richardson, M. Thomson, D. Infield, A high-resolution domestic building occupancy model for energy demand simulations, *Energy Build.* 40 (2008) 1560–1566. <https://doi.org/10.1016/j.enbuild.2008.02.006>.
- [15] A. Sekar, E. Williams, R. Chen, Changes in Time Use and Their Effect on Energy Consumption in the United States, *Joule.* 2 (2018) 521–536. <https://doi.org/10.1016/j.joule.2018.01.003>.
- [16] J. Widén, M. Lundh, I. Vassileva, E. Dahlquist, K. Ellegård, E. Wäckelgård, Constructing load profiles for household electricity and hot water from time-use data-Modelling approach and validation, *Energy Build.* 41 (2009) 753–768. <https://doi.org/10.1016/j.enbuild.2009.02.013>.
- [17] R. Baetens, R. De Coninck, J. Van Roy, B. Verbruggen, J. Driesen, L. Helsen, D. Saelens, Assessing electrical bottlenecks at feeder level for residential net zero-energy buildings by integrated system simulation, *Appl. Energy.* 96 (2012) 74–83. <https://doi.org/10.1016/j.apenergy.2011.12.098>.
- [18] J. Keirstead, A. Sivakumar, Using Activity-Based Modeling to Simulate Urban Resource Demands at High Spatial and Temporal Resolutions, *J. Ind. Ecol.* 16 (2012) 889–900. <https://doi.org/10.1111/j.1530-9290.2012.00486.x>.
- [19] J. Widén, A. Nilsson, E.W.-E. and Buildings, undefined 2009, A combined Markov-chain and bottom-up approach to modelling of domestic lighting demand, Elsevier. (n.d.). <https://www.sciencedirect.com/science/article/pii/S0378778809000978> (accessed January 24, 2020).
- [20] M. López-Rodríguez, I. Santiago, D.T.-M.-E. Policy, undefined 2013, Analysis and modeling of active occupancy of the residential sector in Spain: An indicator of residential electricity consumption, Elsevier. (n.d.). <https://www.sciencedirect.com/science/article/pii/S0301421513007386> (accessed January 24, 2020).
- [21] M.E.J. Newman, The structure and function of complex networks, *SIAM Rev.* 45 (2003) 167–256. <https://doi.org/10.1137/S003614450342480>.
- [22] S.H. Strogatz, Exploring complex networks, *Nature.* 410 (2001) 268–276. <https://doi.org/10.1038/35065725>.
- [23] D.J. Watts, S.H. Strogatz, Collective dynamics of “small-world” networks, in: *Struct. Dyn. Networks*, Princeton University Press, 2011: pp. 301–303. <https://doi.org/10.1038/30918>.
- [24] D.J. Watts, A simple model of global cascades on random networks, *Proc. Natl. Acad. Sci. U. S. A.* 99 (2002) 5766–5771. <https://doi.org/10.1073/pnas.082090499>.
- [25] S.S. Shen-Orr, R. Milo, S. Mangan, U. Alon, Network motifs in the transcriptional regulation network of *Escherichia coli*, *Nat. Genet.* 31 (2002) 64–68. <https://doi.org/10.1038/ng881>.

- [26] I. Bison, Sequence as Network: An Attempt to Apply Network Analysis to Sequence Analysis, in: 2014: pp. 231–248. https://doi.org/10.1007/978-3-319-04969-4_12.
- [27] A. Hagberg, P. Swart, D. Chult, Exploring network structure, dynamics, and function using NetworkX, (2008). <https://www.osti.gov/biblio/960616> (accessed January 28, 2020).