

Systematicity over the course of early development: an analysis of phonological networks

Catherine E. Laing¹

¹ University of York, York, UK

Author Note

All code and associated data for this manuscript can be found at <https://github.com/cathelaing/NetworkGraphs>. Data provided in the PhonBank corpus was collected through support from grant number NIH-NICHHD RO1-HD051698.

Correspondence concerning this article should be addressed to Catherine E. Laing, Department of Language and Linguistic Science, University of York, Heslington, YO10 5DD. E-mail: catherine.laing@york.ac.uk

Abstract

This paper explores the early lexicons of nine infants acquiring English or French to determine the extent of systematicity in the early vocabulary, and how this changes over time. Network graphs are generated from the point of first word production in the data set until age 30 months. Two measures of systematicity - mean path length and clustering coefficient - are analysed to establish the extent to which the early productive lexicon consists of closely-connected clusters of similar-sounding forms. Results show that early production is highly systematic when compared to random, prototypical and adult phonological networks, but that the network becomes more dispersed as it increases in size. Connectivity within the network is consistently higher for infants' actual productions when compared with the adult target forms, and this effect increases over time. This suggests a systematic approach to production over the course of early development.

Keywords: systematicity, phonological development, networks analysis

Systematicity over the course of early development: an analysis of phonological networks


Infants' early words are phonologically similar to one another, if not in the vocabulary items they choose to produce, but in the way they produce them. This has been well-documented in a number of previous studies (e.g. Szreder, 2013; Vihman, 2016; Waterson, 1971), and can be clearly observed in datasets of early productions. For example, an inspection of Deuchar and Quay's (2000) record of their Spanish-English bilingual child's vocabulary acquisition shows that many of her earliest words are produced with an open CV syllable, and she produces a number of identical forms to refer to a range of different (though phonologically-similar) words. This suggests that infants may be drawing on a systematic approach to early productions, whereby a small subset of simple phonological forms are used to produce a range of more varied and phonologically-challenging adult targets. We would expect, therefore, that newly-acquired productions 'cluster together' with existing forms, with high phonological similarity between new words and existing words in the lexicon. This paper will test this by analysing network graphs of the early vocabulary to determine how similar infants' words are to one another, and how this changes over time.

Network analysis is an increasingly popular method of analysing lexical acquisition, and offers an opportunity to consider production data on a larger scale than has previously been possible. Network models allow the analysis of connectivity within a system (in our case, the lexical or phonological system), and can track how that connectivity changes over time. In the case of language development, the nodes (individual items) in the network typically consist of words, and these are connected (or not) depending on how similar two nodes are in phonological or semantic space. Two words that are more similar to one another will be positioned closer together in the network, and less similar nodes further apart. Because this is a convenient way to think about language development over time, a number of studies have considered vocabulary acquisition within this framework. Analyses have considered infants acquiring their first language (e.g. Amatuni & Bergelson, 2017; Fourtassi, Bian, & Frank, 2020) and adults acquiring a novel language (e.g. Luef, 2022; Mak & Twitchell, 2020; Siew & Vitevitch, 2020).

Systematicity is apparent across many linguistic domains, from phonology to syntax, and is noted as being pervasive in linguistic structure (Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015; Nölle, Staib, Fusaroli, & Tylén, 2018). It is suggested that systematicity may have evolved through transmission over generations, as linguistic structure becomes increasingly ordered through use (Kirby, Cornish, & Smith, 2008). This may be a “design feature” of language that makes it easier to acquire and transmit, as has been shown for adults (Kirby et al., 2008; Raviv, Heer Kloots, & Meyer, 2021), children (Raviv & Arnon, 2018), and in computational research (Monaghan, 2011). Systematicity may thus be important in supporting the cognitive processes required to acquire a first language. In early phonological development, systematicity is most comprehensively discussed in work by Vihman and colleagues (e.g. Vihman, 2016, 2019; Vihman & Croft, 2007; Vihman & Keren-Portnoy, 2013). In more than four decades of work, incorporating data from a large number of infants acquiring an impressive range of languages, Vihman demonstrates a clear systematicity in infants’ path to target-like word production. From the initial ‘surprisingly accurate’ forms that appear in the first stages of word production, infants are shown to draw on what they know: they generally choose, for first production, words that are simple in their target phonological form, with consonants that are already familiar from the most common syllables of canonical babble (McCune & Vihman, 2001). These forms are, in Vihman’s terms, *selected* for first word production owing to their easily-producible features. As the vocabulary grows, **infant** must necessarily acquire forms that do not contain such accessible phonological or segmental properties. Here we begin to see regression in the accuracy of early production, as infants systematically adapt forms to fit the most common structures and segments in their repertoire. When words are systematically altered to fit a dominant pattern in the child’s output, these forms are said to be *adapted*. Systematicity is apparent not just in the earliest words, but across the trajectory of acquisition as infants deal with the challenges of early word production by relying on well-rehearsed output forms. Over the first months of lexical development at least, infants’ productions of newly-acquired words are likely to match their productions of existing words in the lexicon.

Analysing infants’ early word productions is highly resource-intensive, and so previous work

has typically drawn on a case study design [e.g. ADD REFS], or analysis of a subset of words from infants' wider lexicons [e.g. ADD REFS], with varying timescales of development targeted [ADD REFS]. Ideally, to fully understand the role that systematicity plays, analyses would incorporate a randomly- or systematically-sampled range of words making up a large proportion of the early vocabulary, observing a wide developmental timescale. Drawing on a detailed, word-by-word analysis of the developing vocabulary would not be feasible, but a networks approach allows us to consider systematicity across a much larger set of words and along a wider developmental trajectory than has typically been drawn upon in research in this area. In recent work (Laing, 2024), I draw on network analysis to analyse systematicity in the developing lexicon. I show that in the first three years of life, infants' production of new words can be predicted based on the words they already produce and how they produce them. That is, a word is more likely to be acquired if it is produced in a way that is phonologically similar to existing words in the productive repertoire, particularly when the new word is similar to a cluster of existing phonologically-similar forms in the output. This effect becomes stronger over time, suggesting that systematicity is more relevant to later word learning (at least, upto age 30 months) than in the first few months of word production. An analysis of target forms showed similar results, though with weaker predictive power. Moreover, the phonological properties of infants' word productions are more similar to one another than their adult target forms would suggest. These findings support the more fine-grained analyses presented by Vihman and colleagues, referenced above.

These methods have already been used to analyse vocabulary networks across different kinds of data. Laing (2024) drew on corpus data of fortnightly word production across nine infants; Kalinowski and colleagues (in prep) take a similar approach to Laing by analysing vocabulary norms from >1000 Norwegian infants taken longitudinally at up to six individual time-points. Their results also show evidence for systematicity in the target words acquired in early acquisition (the analysis of vocabulary norms means it is not possible to observe infant productions of these words ). Systematicity was also identified in early word learning by Siew and Vitevitch (2020), in an analysis of vocabulary norms of children aged 3-9 years acquiring English and Dutch. Fourtassi and

colleagues (2020), on the other hand, found no support for systematicity in their analysis of infants acquiring a range of 10 different languages, instead showing that salient properties of the input, rather than previously-learned phonological properties, predicted learning.

Note that the previous studies reported above (with the exception of Laing (2024)) all draw on vocabulary norming data, meaning that infants' actual productions (i.e. the way they produce words) is not considered. One of the key strengths of Laing (2024) and the current paper is the application of phonological network analysis to real production data. From decades of work on early phonological development, we know that infants' earliest words are often far-removed, phonologically speaking, from their adult targets. For example Priestley's (1977) son is reported to produce *banana* as /bajan/, *sucker* as /fajak/, *chocolate* as /kajak/ and *medicine* as /mejas/ within the same week at age 1;10. The adult target forms are all highly variable in form, while the child forms all share the systematic implementation of a disyllabic production pattern with medial /j/, following the structure CVCVC. Such systematicity cannot even be hinted at from an analysis of the target forms only, thereby losing what may be a crucial aspect of the acquisition process. Moreover, Fourtassi and colleagues (2020) and Siew and Vitevitch (2020) draw on vocabulary norming data from a cross-sectional sample of infants, meaning it is only possible to take a very general view of acquisition, with no scope for considering individual variability across the sample. This work thus cannot address any questions about systematicity in early acquisition (though note that this was not the intention of either paper).

This paper contributes a novel approach to the study of developmental vocabulary networks by analysing network *graphs*, rather than network *growth models* to study phonological networks in the developing vocabulary. All the previous work reported above uses network growth algorithms to test whether learning can be predicted based on the words that infants already know or produce. Such an approach can be used to test and/or compare different theoretical models of acquisition, whereby different network growth algorithms that index different predictions can be used as predictors in statistical models. Moreover, network growth models analyse connectivity (are two words similar, yes or no? if yes then they are connected in the network), rather than phonological

distance (*how* similar are two words in the network?), and analyse the possibility of a new word being added to the network at the next time-point, rather than the static properties of a network at a given time-point. Network graphs, on the other hand, allow us to understand more about the properties of the network at a given time-point: how ordered/random the network is, how dense its clusters are, and how closely connected words are to one another across the network. Crucially, this will allow us to make predictions about network properties over time that reflect word selection and adaptation, and thus presents an opportunity to apply Vihman's framework on a broader scale. Following Laing (2024), this paper will analyse infants' actual productions as well as the adult target forms, to test the extent to which systematicity is present in production, in terms of both the words infants select in early development, and how they produce them.

Research questions and predictions

This paper builds on previous network analyses by drawing on network graphs, instead of growth algorithms, to look more closely at the phonological distance (that is,) between individual words in the developing lexicon. In doing so, it attempts to address the following questions:

1. How systematic are early word productions (both actual and target), and (how) does this change over time?
2. Are the phases of word selection and adaptation identifiable in the dataset?

To test these questions, network graphs will be generated using the *igraph()* package (Csardi & Nepusz, 2006) in R (R Core Team, 2020) for both the actual and target data. To address the first question, properties of the graphs will be analysed to determine 1) how closely connected individual words are to one another; 2) how dense the overall distribution of words is in the network; and 3) how/whether this changes over time. Following Vihman's work, and findings presented by Laing (2024), it is expected that the early vocabulary will become increasingly systematic over time. This would be reflected in denser clusters of phonologically-similar forms and shorter distance between words. Following previous research [REFS], this should be true for both actual and target data.

Simulated networks will be used to compare the real networks against both highly systematic and random networks to determine the extent of systematicity present in the data, and developmental changes over time.

To address the second question, network graphs of infants' actual productions will be compared with those of the target form, to trace the 'target-likeness' of individual productions, and how this changes over time. Following Vihman once again, early word selection would be reflected in early similarities between Actual and Target network properties, as target forms are selected to match the structures and segments that infants are able to produce, meaning they should be produced with relative accuracy. Over time, Actual and Target forms are expected to diverge, such that Actual forms show more systematicity in the data than Target forms. Approaches used to test these predictions are outlined in detail below.

Methods

Data extraction and preparation

This study draws on the same data as that analysed by Laing (2024). This was drawn from two corpora on PhonBank (Rose & MacWhinney, 2014): Providence (American English - Demuth, Culbertson, & Alter, 2006) and Lyon (French - Demuth & Tremblay, 2008). These were selected due to their equivalent data collection methods and the fact that the infants' productions, as well as the corresponding target forms, are phonetically transcribed. Nine infants' (5 English, 4 French, 4 boys overall) data were extracted using Phon (Hedlund & Rose, 2020), from the transcript with their first-recorded word to the final transcript taken at age 2;6. Infants were recorded in the home on a fortnightly basis, participating in naturalistic interactions with their caregivers. Two of the American infants were recorded weekly during some periods of data collection, but this is not an issue for this analysis since no between-child comparisons will be made. See Demuth et al. (2006) and Demuth and Tremblay (2008) for full details of data collection.


```

types.corpus <- comparison_data %>%
  group_by(corpus, Speaker) %>%
  distinct(Gloss, .keep_all=T) %>%
  tally() %>% summarise(n=sum(n))

types <- comparison_data %>%
  group_by(Speaker) %>%
  distinct(Gloss, .keep_all=T) %>%
  tally() %>% summarise(n=sum(n))

tokens <- comparison_data %>%
  group_by(Gloss, Speaker, Session) %>%
  tally() %>% ungroup() %>%
  summarise(mean_tok=mean(n),
            sd_tok=sd(n))

tokens.corpus <- comparison_data %>%
  group_by(Gloss, Speaker, Session, corpus) %>%
  tally() %>% ungroup() %>%
  group_by(corpus) %>%
  summarise(mean_tok=mean(n),
            sd_tok=sd(n))

```

Extracted data was filtered to include only words featuring on the communicative development inventory (CDI) of the respective language, including all variants of a given form (including a total of 680 possible word types for American English and 690 for European French). For example, plurals were included alongside the singular noun form, and variable verb conjugations

alongside their corresponding infinitive verb form. There were 5483 non-CDI words in the initial dataset, all of which were excluded from the analysis (by language: 2224 in French and 3259 in English). The final dataset includes 3320 word types overall, aggregated across infants (English=2024, French=1296).

Overall 58.63% of words in the dataset were excluded due to not being on the CDI (CHECK THIS). While filtering the data in this way makes the analysis easier to compare across similar studies that also use CDI vocabulary measures (and provides a more manageable dataset for the analysis, given the computational load of comparing all words with all other words in the data), the loss of so many words from the infants' vocabularies means results likely won't capture all relevant aspects of the infants' early production. These data were used to create a network for each infant for each month of data, whereby all words in a given month that hadn't been produced in previous months were included as new items in the network. When a word had been produced in previous months, it was not included. While this means that the data does not capture change in the production of a single form over time, it allows us to observe network growth at the point of acquisition for each word form.

To determine the structure of the network, the first step was to create distance values between each word and each other word in the network. This was first done using a 'global' network of all forms produced by the infant up to the final session at 2;6, to create a large distance matrix for each infant that incorporated all word productions. Essentially, this global network reflects the distance between every word and every other word in each child's productive vocabulary at 2;6. Distance values were established using methods set out in Monaghan et al. (2010), using distinctive features to generate a set of phonetic values for each word that could then be compared with all other words (note that only consonants were analysed, given that vowels are highly variable in early production and also very difficult to transcribe accurately, Donegan, 2013; Kent & Rountrey, 2020). Euclidean distance between the values of each word and each other word in each infant's global network was then used to determine how close/distant words were from one another. By this measure, word pairs with a distance of 0 have the same consonants produced in the same word position (but may differ

in vowels), such as *bat* and *bet*. Often, infants produced multiple tokens of the same word type in a given month, often with variability in the way that different tokens were realised. Because it was not possible to generate networks with all word tokens included (even with only single word types included, the final dataset for all nine infants includes over 5.5 million data points, once distance between each word and each other word is calculated), a mean value for each distinctive feature was established across tokens, meaning that each word's distinctive feature value represents the variability of the infant's production of a given word. For example, if an infant produced two tokens of the word *doggie* as [dagi] and [dati], respectively, each of the distinctive feature values for /g/ and /t/ would be averaged across tokens to create an "average production" of that word. This may not be a perfect measure, but it is more representative than taking, for example, the first instance of each word type.

Distance scores were generated between each word and each other word in each child's dataset, for both Target and Actual forms. These scores were then normalised, and a normalised distance of 0.25 was chosen to indicate connectivity. That is, words were said to be connected in the network if their distance score was 0.25 or less. This accounted for the lower quartile of connectivity across the dataset. For a full overview of the data preparation process, including validation of the 0.25 connectivity threshold, see Laing (2024, supplemental materials).

The final dataset includes 3320 word types in total, with 2024 in the English corpus and 1296 in the French corpus. On average, infants produced 8 tokens of each word type in a single session (SD = 18; mean English tokens = 6, SD = 11; mean French tokens = 13, SD = 29).

Data analysis

Network graphs. The prepared data was then used to generate a series of network graphs for each infant (for both Target and Actual data) using the *igraph()* package in R (Csardi & Nepusz, 2006). One network was generated per month, for each month in the dataset, based on all words produced in the given month and all months prior. The network at time-point $n+1$ thus included all

word types produced upto and including time-point n , plus all additional words produced for the first time at $n+1$. Monthly network size for each infant is shown in Table 1. The *igraph()* package generates graphs that include all nodes (whether or not they are connected to other nodes¹), and measures the distance between all connected nodes, as well as the clustering of nodes in graphical space. Two example networks are shown in Figure 1, where differences between Actual and Target networks, as well as phonological distance between nodes, and un-connected “hermit” nodes, are visualized. IPA transcriptions of the Target and Actual words in each network are shown in Table 2.

Two key variables will be explored through an analysis of network graphs: *mean path length* and *average clustering coefficient*. These have been calculated from a measure I refer to here as phonological distance, which is derived from the Euclidean distance between pairs of consonants across any given two words in the dataset. Path length is a measure of distance between nodes, and mean path length indexes the average phonological distance (of all connected nodes) within a network; by this measure, we would expect that systematicity in early phonological development would be reflected in *low* mean path length, as words are, on average, more closely connected. Clustering coefficient is an indication of network density: a *higher* average density of nodes in the network indicates denser clusters of similar forms; again, this is what we would expect to see in a network of early phonological development. See Goldbeck (2013) for a full overview of network structures and measures.

Simulated networks. Networks with high phonological systematicity should exhibit properties of prototypical “small-world” network growth, namely a low mean path length and a high average clustering coefficient (Amaral, Scala, Barthelemy, & Stanley, 2011; Steyvers & Tenenbaum, 2005; Watts & Strogatz, 1998). Words should be more densely connected, with shorter connections between words. As this is what we expect to see in the phonological networks tested here, small-world networks serve as a suitable comparison with the real data. If high systematicity is present in the data, then there should be no statistical difference between the real network and a

¹ Recall that any two nodes that have a scaled phonological distance of $>.25$ will not be connected.

small-world network of equal size. To test this, mean path length and clustering coefficient values were generated for both the Target and Actual networks, as described above. Data were then compared to the growth of a simulated small-world network of equal size, known as a Watts-Strogatz network (Watts & Strogatz, 1998). This network was generated computationally in R using the *igraph()* package (Csardi & Nepusz, 2006), and was matched for network size and mean connectivity within the network at each month. The real data were also compared to a similarly-sized but randomly-generated network known as a Erdős–Rényi model. If the real network grows in a systematic way, then we would expect the real data to differ significantly from the randomly-generated Erdős–Rényi network. Again, this was generated using the *igraph()* package (Csardi & Nepusz, 2006), and was matched for network size at each month (but not mean connectivity, as this is not required for a random graph). To run these analyses, mean path length and clustering coefficient were calculated for each monthly graph - the Real data, and the two kinds of simulated data (prototypical systematic network vs. random network) - and the two kinds of Simulated network were tested against the Real network.


Statistical models. Two different analysis methods will be drawn upon to test the two research questions. To test RQ1, network graphs will be compared to simulated small world and random networks of equivalent size to determine whether the data (the *Real* network, drawing from infants' Actual productions) differs from the *Simulated* networks using linear mixed effects regression models. To test RQ2, generalised additive mixed effects models (GAMMs) will be used, since these allow the analysis of non-linear change over time, and can account for statistical differences between two non-linear trajectories of data that may differ in non-linear ways (Sóskuthy, 2017; Wieling, 2018). Fixed effects in the model can include parametric terms, as is typical in regression modelling, and also *smooth terms*, or non-linear fixed effects. Much like mixed-effects linear models, GAMMs can account for random effects in the data; in this case by-subject random effects were included through the addition of *random smooths* in the model.

To account for the fact that adjacent values (i.e. connectivity at month n and month $n+1$) are likely correlated, GAMM modelling includes an autocorrelation parameter; see Sóskuthy (2017)

and Wieling (2018) for full details. Additionally, the start point for each infant's data (i.e. their first recording session) was indexed in the model.

Results

RQ1: How systematic are early word productions, and (how) does this change over time?

 For both mean path length and clustering coefficient, we would expect the Real data to differ significantly from the Simulated Erdős–Rényi (random) network, and for the Real network to show similar properties to the Simulated Watts-Strogatz (small-world) network. Note that the extent of the expected statistical difference is not easy to predict here: if the Real network is very similar to the Watts-Strogatz network then no statistical difference would be expected, but this relies on the Real data being highly systematic, which may not be realistic. In order to fully understand the nature of the data, figures and model outputs will be inspected in relation to these predictions.

The two measures will be discussed in turn. Models include mean path length or average clustering coefficient as the dependent variable, respectively, each with Data type (Real vs. small-world vs. random), Corpus (English vs. French) and Age as fixed effects, and Subject as a random effect with a by-Subject random slope for the effect of age. Initial model comparisons showed that including Network size, alongside Age, improved fit in the model testing clustering coefficient, but not mean path length (see Supplementary materials, S1, for model comparisons). Network size is thus included as a fixed effect in the clustering coefficient model only.

Mean Path Length. Nested model comparisons revealed a significant effect for Data type on mean path length. See Table 3. As shown in Table 4, the Real data had a significantly lower mean path length than the random Simulated data, as predicted. The difference between the Real data and the Simulated small world data was also significant, but the extent of this difference was smaller than for the comparison with the random network, thereby lending support to initial predictions. The extent of these effects is shown clearly in Figure 2, where the comparison between the Real data and the random Simulated data is visually much wider than that of the Real

vs. Simulated small-world network until the very end of the data collection period. Contrary to predictions, there was no change in systematicity over time (i.e. no effect of Age was observed); there was also no effect for corpus. Plots showing the effect of age on the two variables can be found in the Supplementary Materials, S2.

Clustering coefficient. The same main outcomes were found when clustering coefficient was tested in the model. See Tables 3 and 4. The Real data had a significantly higher clustering coefficient than the random Simulated data, but this was significantly lower than that of the small-world Simulated data. Again, the magnitude of the difference was much larger in the Real vs. random Simulated comparison than the Real vs. small world Simulated comparison, again supporting predictions. This is visualised clearly in Figure 3. Again there was no effect for Corpus on the data, but this time there was a significant effect for Age: clustering coefficient values increased over time, suggesting that systematicity became stronger in the data over time; with each additional month, clustering coefficient increased by 0.20%. However, a significant effect for Network size contradicts this; with each additional word added to the network, clustering coefficient decreased by 0%.

Actual vs. Target data. The differences between the Real data and the small-world Simulated data are difficult to interpret, given that there is no clear model of what phonological systematicity would look like in a highly systematic small-world network. To further interrogate systematicity within the data, network properties of the Real (Actual) data analysed above were compared to the Real Target data - that is, the phonological distance (taken from the IPA transcriptions) between the child's Actual production and its Target counterpart was analysed. Target data serves as an appropriate proxy for connectivity and clustering within a "standard" phonological network, albeit a network that is constrained by words produced in early acquisition, and further constrained by the fact that only a subset of these (i.e. CDI words) are included in the data set. We would expect that mean path length and clustering coefficient would each show higher

systematicity in the Actual network than the Target network². Basic model structure was the same as reported above, but with only Real data (Actual vs. Target) considered as a fixed effect. The inclusion of Network size as a fixed effect improved model fit for both dependent variables and so was included in both models.

There was a significant effect for Data type on mean path length. See Table 3. Model outputs revealed that Target data had a significantly higher mean path length than Actual data (see Table 5 and Figure 4). This time there was a significant effect for Corpus, whereby the French data had a higher mean path length than the English data overall. Network size, but not Age, significantly affected mean path length, which increased as Network size increased, indicating a decrease in systematicity over time: with each new word added to the network, mean path length decreased by 0%.

The effect of Data type on clustering coefficient was also significant. Average clustering coefficient was significantly lower in Target compared with Actual data. See Figure 5. The French data had a significantly lower mean clustering coefficient than the English data. Again, there was a significant effect for Network size, and this was consistent with the result reported above for mean path length, in that it indicated a decrease in systematicity with increased network size. As new words were acquired in the network, average clustering coefficient decreased by 0.04%. Again, Age was not a significant predictor of changes in clustering coefficient.

```
SWD_red$start.event <- SWD_red$session_ordinal == 1

globalthresholds$start.event <- globalthresholds$threshold == 0.01


# create dummy variable for actual vs. target
```

² Note that this result was observed for network growth models in Laing (2024), whereby the Actual network was found to be a better predictor of learning based on the known network of each child.


```
globalthresholds$IsActual <- (globalthresholds$data_type == "actual")*1
```

RQ2. Is there evidence of word selection and adaptation in the dataset?

To address the second research question, the phonological distance between Target and Actual forms was taken as a proxy of word selection and adaptation. That is, if a word is produced in a target-like way (i.e. assumed to be selected³), then the phonological distance between the Target form and the way it is produced (Actual form) should be low. The opposite is true for adapted forms, as we expect, by definition, a non-target-like production and thus a higher distance between Target and Actual form. This measure is not perfect, but coding selected/adapted forms would otherwise have to be done by hand, which is not feasible across such a large dataset. Following Vihman's (2019) framework, we would expect low distance between Actual and Target forms earlier on in development as words are selected, and higher distance later as word adaptation begins to take hold.

GAMMs were used to examine connectivity of the infants' Actual and Target networks and how these changed over time. These were run using the *mgcv()* package in R (Wood, 2011). These models analyse the extent to which the two networks differ (or not) from one another across infants, and how this changes non-linearly month-by-month. The model tested mean number of connections in the network (average number of connections of each node in the network, or mean k) as the dependent variable, working on the assumption that connectivity in the Target vs. Actual networks would be similar during periods of word selection (i.e. Actual and Target words are similar to one another and so distribution of connectivity should be similar), and would differ during periods of adaptation. Specifically, periods of adaptation should lead to higher connectivity in the Actual network than the Target network, since we expect productions to be more similar (and thus more well-connected) in Actual forms; we would expect connectivity across data types to diverge at the point that word adaptation begins to take hold  higher number of connections (higher mean k) for

³ though note that, while a selected form is, by definition, target-like in phonological form, a target-like form isn't necessarily a selected form.

Actual vs. Target networks is expected during periods of adaptation, and no difference in connectivity is expected for periods of selection. Data type (Actual vs. Target) and Corpus (English vs. French) were included as parametric terms, with Data type being the variable of interest in the model. Network size and Age were included as smooth terms, as well as by-Subject and by-Data type random smooths for the effect of Age, which account for by-Subject and by-Data type differences in the data over time. To test for an effect of Data type, model comparisons were run using the *compareML()* function from the *itsadug()* package (Rij, Wieling, Baayen, & Rijn, 2022): the full model including the effect of Data type and the by-Data type random smooth was compared to a model without these terms. Because model summaries for GAMM smooths may be non-conservative (Sóskuthy, 2017), smooth plots will be observed alongside any significant effects to determine relevant trends in the data.

Model comparisons revealed a significant effect for Data type on connectivity in the networks over time. See Table 3. Figure 6 shows the difference between Actual and Target connectivity over the course of development. The red line indicates periods of significant difference, showing that Actual vs. Target connectivity was significantly different throughout the period of analysis; Actual forms were always more well-connected than Target forms. This contrasts with the expectations set out above. However, Figure 6 clearly shows an increase in the difference in connectivity between Actual and Target forms over the period of data collection, supporting the expectation that Target and Actual forms are more similar earlier on in development, with an increasing difference in mean connectivity over the course of the analysis, favouring higher connectivity in Actual, compared with Target, forms.

Discussion

This paper set out to test the presence of systematicity in infants' developing lexicons by analysing phonological network graphs from nine infants acquiring French or English. Network graphs allow a close-up view of phonological similarity between forms within the network (via mean path length) and the extent to which groups of phonologically similar forms cluster together (via

average clustering coefficient). Systematicity in the network would be reflected in shorter distance between forms and higher clusters of phonologically similar forms. The analysis also sought to identify periods of selection and, later, adaption - indicating a shift towards increasing systematization - in the data.

The first research question asked whether or not systematicity could be identified using a network graphs analysis, and, if so, whether or not this changed over time. This essentially presents a replication of analyses on the same data by Laing (2024), using a different analytical approach. Comparisons of network graphs generated using the real data against random and highly systematic simulated network graphs lend support towards the presence of systematicity within the data, and this was strengthened with a follow-up analysis comparing networks of the infants' actual productions with those of the target forms. Overall, infants' early productions had a shorter mean path length and formed denser clusters of similar forms within the networks (i.e. higher average clustering coefficient) than simulated random networks and networks of the target phonological forms, though these were less systematic than prototypical highly systematic "small world" simulated networks. These findings support those of Laing (2024) to suggest that systematicity is present in early phonological productions.

The picture gets a little more complex when changes over time are considered. Laing (2024) identified an increase in the predictive power of the network model over time, which indicated increasing systematicity in the network. However, this was not the case for the present analysis. Looking at the comparison between Actual and Target data, neither mean path length nor clustering coefficient changed with age. However, network size was a predictor of both effects, but in the unexpected direction: higher mean path length and lower clustering coefficient were observed as network size increased. This is contrary to what was found by Laing (2024) and appears to indicate a decrease in systematicity over time. These contrasting findings are likely driven by differences in what was being analysed: the network growth algorithms in Laing's (2024) paper predict the likelihood of a given word being added to the network in the next month, and, over time, the network was more likely to acquire new words that would connect to the most densely-connected

words in the existing network (controlling for network size). Adding the present findings to this picture, the densest clusters at later timepoints were in fact less dense than those at earlier timepoints, with larger phonological distances between connected words. This aligns with what we know about systematicity over the course of early phonological development, as the kinds of words being targeted for production become more variable. This is demonstrated in case study accounts of infants' early words, where we see the establishment of different production patterns, or templates (Vihman, 2019) over time. This is clearly demonstrated in Waterson's (1971) case study of her son's production at 18 months, ~~for example, where~~ five distinct structures are identified in his data, to which newly-acquired words are systematically adapted. In this example, we see systematicity becoming more prevalent in the data (as a wider range of templates gives rise to more opportunity for word adaptation) but clusters of similar words may be less dense, as adaptation takes place in a number of different - but systematic - ways.

The second research question attempted to identify periods of word selection and adaption in the data using generalised additive mixed models (GAMMs). This analysis worked on the assumption that overall connectivity in the network (mean number of connections per node, or mean k) should be similar for Actual and Target forms during periods of word selection, since infant productions should be more phonologically accurate for selected words, and thus similar in connectivity to the target forms. Periods of adaptation, on the other hand, should see a difference in connectivity between Target and Actual data; new Target forms in the network will be more distant, and less likely to connect to existing forms, whereas Actual forms should continue to connect to existing words in the network. Thus, we expect a divergence between connectivity in the Target and Actual networks. This was, to some extent, borne out in the data, though distinct periods of selection and adaptation could not be identified. Instead, a gradual increase in the difference in connectivity between Actual and Target networks was observed over time; the difference in connectivity was always significant, and mean k was always higher in Actual than Target networks.

One key question arising from these findings is the extent to which network graphs - in particular the measures used in this study - can help us understand systematicity. Even though this

analysis incorporated a more close-up analysis of network properties than those that draw on network growth models alone - i.e., by taking into account distribution of the nodes within the networks across three measures (mean path length, average clustering coefficient, and mean k) - still the analyses ~~all~~ abstract away from the detail of early word production, and specifically what drives connectivity and clustering within the network. The measures used to generate phonological distance were not able to take into account holistic or prosodic properties of early words, so we lose a potential source of connectivity within the networks (for example, a reliance on consonant harmony, which reveals within-word, as well as potential between-word, systematicity). alinowski and colleagues (in prep) take steps towards addressing this in their analysis, showing the potential for even more nuanced indices of phonological distance by iterating across, as well as between, similarities within words in the network. Moreover, infants are likely to have drawn on different approaches to early word production, some which might have been represented as systematic more effectively by the measures used here. Finally, the focus on mean k to analyse word selection and adaptation is a crude measure for identifying changes in the data that are likely very subtle. Future work may want to draw on cluster analysis to observe these changes in the data more closely. That being said, phonological distance between Target and Actual forms may be a useful measure for objectively identifying word selection or adaptation in future studies.

While this analysis presents a more nuanced view of systematicity than in previous studies that draw entirely on network growth models, including Laing (2024), Fourtassi and colleagues (2020) and Siew and Vitevitch (2020), closer inspection of the network graphs themselves may have been useful in supporting and explaining the findings in more detail. Future work could combine computational analyses of networks with a more impressionistic analysis of infants' early word productions to bring together these two very different but equally valuable methodologies. Indeed, drawing on modelling to understand large-scale data makes it possible to quantify findings in a more rigorous way, but it abstracts away for the nature of early productions and means we miss out on the detail of the patterns that are being drawn upon. It also overlooks the extent to which infants take very individualised paths in phonological development (Vihman, 1993), and the possibility that

broad-scale generalization is simply not reflective of the reality of early production. Future studies in this area may want to consider how variability between infants, and infants acquiring different languages, is represented in vocabulary growth networks.

Overall, these findings support a case for systematicity in early development. The analysis of network graphs supports and builds on existing studies in this area - those that present a 'close-up' case-study analysis (e.g. Szreder, 2013; Waterson, 1971) and those that draw on computational methods to analyse large data sets in a generalised way (e.g. Fourtassi et al., 2020; Laing, 2024) - to present a nuanced evaluation of a large-scale early production data set. Findings suggest that the developing network is characterised by an increasing number of dense clusters of similar-sounding words, and that systematicity is present in early production from the beginning.

References

```
r_refs(file="r-references.bib")
```

Amaral, L. A. N., Scala, A., Barthelemy, M., & Stanley, H. E. (2011). Classes of small-world networks. In *The structure and dynamics of networks* (pp. 207–210). Princeton, NJ: Princeton University Press.

Amatuni, A., & Bergelson, E. (2017). *Semantic Networks Generated from Early Linguistic Input* [Preprint]. <https://doi.org/10.1101/157701>

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695. Retrieved from <https://igraph.org>

Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of english. *Language and Speech*, 49(2), 137–174.

Demuth, K., & Tremblay, A. (2008). Prosodically-conditioned variability in children's production of french determiners. *Journal of Child Language*, 35(1).

Deuchar, M., & Quay, S. (2000). *Bilingual acquisition: Theoretical implications of a case study*. Oxford, UK: Oxford University Press.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences*, 19(10), 603–615. <https://doi.org/10.1016/j.tics.2015.07.013>

Donegan, P. (2013). Normal vowel development. In M. J. Ball & F. E. Gibbons (Eds.), *Handbook of vowels and vowel disorders* (pp. 24–60). Psychology Press.

Fourtassi, A., Bian, Y., & Frank, M. C. (2020). The growth of children's semantic and phonological networks: Insight from 10 languages. *Cognitive Science*, 44(7), e12847. <https://doi.org/10.1111/cogs.12847>

Golbeck, J. (2013). *Analyzing the Social Web*. San Francisco, UNITED STATES: Elsevier Science & Technology.

- Hedlund, G., & Rose, Y. (2020). *Phon 3.1 [computer software]*. Retrieved from <https://phon.ca>
- Kalinowski, J., Hansel, L., Vystrčilová, M., Ecker, A., & Mani, N. (in prep). *The development of early phonological networks: An analysis of individual longitudinal vocabulary growth*.
- Kent, R. D., & Rountrey, C. (2020). What acoustic studies tell us about vowels in developing and disordered speech. *American Journal of Speech-Language Pathology*, 29(3), 1749–1778. https://doi.org/10.1044/2020_AJSLP-19-00178
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31), 10681–10686. <https://doi.org/10.1073/pnas.0707835105>
- Laing, C. E. (2024). Phonological networks and systematicity in early lexical acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <https://doi.org/https://doi.org/10.1037/xlm0001368>
- Luef, E. M. (2022). Growth algorithms in the phonological networks of second language learners: A replication of Siew and Vitevitch (2020a). *Journal of Experimental Psychology: General*, 151(12), e26–e44. <https://doi.org/10.1037/xge0001248>
- Mak, M. H. C., & Twitchell, H. (2020). Evidence for preferential attachment: Words that are more well connected in semantic networks are better at acquiring new links in paired-associate learning. *Psychonomic Bulletin & Review*, 27(5), 1059–1069. <https://doi.org/10.3758/s13423-020-01773-0>
- McCune, L., & Vihman, M. M. (2001). Early phonetic and lexical development. *Journal of Speech, Language, and Hearing Research*, 44(3), 670–684. [https://doi.org/10.1044/1092-4388\(2001/054\)](https://doi.org/10.1044/1092-4388(2001/054))
- Monaghan, P. (2011). Design features of language emerge from general-purpose learning mechanisms. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 33, 2661–2666. <https://doi.org/https://escholarship.org/uc/item/7q1736rb>
- Monaghan, P., Christiansen, M. H., Farmer, T. A., & Fitneva, S. A. (2010). Measures of

- phonological typicality. *The Mental Lexicon*, 5(3), 281–299.
<https://doi.org/10.1075/ml.5.3.02mon>
- Nölle, J., Staib, M., Fusaroli, R., & Tylén, K. (2018). The emergence of systematicity: How environmental and communicative factors shape a novel communication system. *Cognition*, 181, 93–104. <https://doi.org/10.1016/j.cognition.2018.08.014>
- Priestly, T. M. S. (1977). One idiosyncratic strategy in the acquisition of phonology. *Journal of Child Language*, 4, 45–65.
- R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Raviv, L., & Arnon, I. (2018). Systematicity, but not compositionality: Examining the emergence of linguistic structure in children and adults using iterated learning. *Cognition*, 181, 160–173. <https://doi.org/10.1016/j.cognition.2018.08.011>
- Raviv, L., Heer Kloots, M. de, & Meyer, A. (2021). What makes a language easy to learn? A preregistered study on how systematic structure and community size affect language learnability. *Cognition*, 210, 104620. <https://doi.org/10.1016/j.cognition.2021.104620>
- Rij, J. van, Wieling, M., Baayen, R. H., & Rijn, H. van. (2022). *itsadug: Interpreting time series and autocorrelated data using GAMMs*.
- Rose, Y., & MacWhinney, B. (2014). The PhonBank Project: Data and software-assisted methods for the study of phonology and phonological development. In J. Durand, U. Gut, & G. Kristoffersen (Eds.), *The Oxford handbook of corpus phonology* (pp. 380–401). Oxford, UK: Oxford University Press.
- Siew, C. S. Q., & Vitevitch, M. S. (2020). An investigation of network growth principles in the phonological language network. *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0000876>
- Sóskuthy, M. (2017). *Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction*. arXiv. Retrieved from <http://arxiv.org/abs/1703.05339>
- Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks:

Statistical analyses and a model of semantic growth. *Cognitive Science*, 29(1), 41–78.

https://doi.org/10.1207/s15516709cog2901_3

Szreder, M. (2013). The acquisition of consonant clusters in polish: A case study. In M. M.

Vihman & T. Keren-Portnoy (Eds.), *The emergence of phonology: Whole-word approaches and cross-linguistic evidence* (pp. 343–361). Cambridge: Cambridge University Press.

<https://doi.org/10.1017/CBO9780511980503.016>

Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics*, 21, 61–82.

Vihman, M. M. (2016). Prosodic structures and templates in bilingual phonological development.

Bilingualism: Language and Cognition, 19(1), 69–88.

<https://doi.org/10.1017/S1366728914000790>

Vihman, M. M. (2019). *Phonological templates in development*. Oxford, UK: Oxford University Press.

Vihman, M. M., & Croft, W. (2007). Phonological development: Toward a “radical” templatic phonology. *Linguistics*, 45(4). <https://doi.org/10.1515/LING.2007.021>

Vihman, M. M., & Keren-Portnoy, T. (2013). The emergence of phonology: Whole-word approaches, cross-linguistic evidence. In M. M. Vihman & T. Keren-Portnoy (Eds.), *The emergence of phonology: Whole-word approaches and cross-linguistic evidence*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511980503.002>

Waterson, N. (1971). Child phonology : A prosodic view. *Journal of Linguistics*, 7(2), 179–211.

<https://doi.org/10.1017/S0022226700002917>

Watts, D. J., & Strogatz, S. H. (1998). *Collective dynamics of “small-world” networks*. 393, 3.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling:

A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, 70, 86–116. <https://doi.org/10.1016/j.wocn.2018.03.002>

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1), 3–36.

Table 1		Number of word types in each child's network at each month. Network size is cumulative such that values in each month include the word types produced in all previous months. Empty cells represent months in which a given infant did not have a recording.														
Speaker	Corpus	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Alex	English	NA	NA	NA	NA	NA	2	6	8	26	42	56	63	94	96	101
Lily	English	NA	NA	12	15	NA	16	20	26	53	117	167	214	272	325	368
Naima	English	6	24	40	62	140	221	274	338	367	406	441	455	475	484	504
Violet	English	NA	NA	NA	3	NA	6	9	17	39	59	109	164	198	224	268
William	English	NA	NA	NA	NA	NA	35	NA	76	99	157	180	200	214	221	235
Anais	French	NA	2	4	9	20	21	46	48	51	54	56	70	120	133	NA
Marie	French	NA	108	110	112	NA	NA	117	123	131	134	137	164	203	234	260
Nathan	French	NA	2	5	NA	6	14	19	21	25	27	32	33	43	57	NA
Tim	French	3	17	21	NA	40	60	80	106	148	173	219	281	304	310	333

Table 2

IPA transcriptions of Actual and Target forms produced by one English-acquiring infant's data, including one token of all words produced upto and including 14 months.

Gloss	Age	IPA _{target}	IPA _{actual}
baby	13	beɪbi	tɛdɛ
cow	13	kaʊ	hʌ
daddy	13	dædi	dɛðɪ
dog	13	dʌg	dʌ
duck	13	dʌk	dɛ
duckie	13	dʌki	dɪðeɪ
hi	13	haɪ	hɛ
mommy	13	mami	ʌwɛ
moon	13	mun	mʌm
puppy	13	pʌpi	bəbɛ
teddy	13	tɛdi	dædɪ
yes	13	jæ	hɛ
ball	14	bʌl	ʊwʌ
doggie	14	dʌgi	hæveɪ
eye	14	aɪ	ɛ
kitty	14	kɪti	kɪzɪ

Table 3

Outputs from nested model comparisons testing the effect of data type (Real vs. Simulated and Actual vs. Target on mean path length and clustering coefficient.

Model	Df	Chisq	p
Mean Path Length (Real vs. Simulated)	2	585.94	<0.001
Mean Path Length (Actual vs. Target)	1	187.86	<0.001
Clustering Coefficient (Real vs. Simulated)	2	1325.23	<0.001
Clustering Coefficient (Actual vs. Target)	1	269.41	<0.001
Mean connectivity (Actual vs. Target): GMM	3	182.63	<0.001

Table 4

Outputs from linear mixed effects regression models testing comparisons of Real vs. Simulated data on mean path length and clustering coefficient.

Effect	Mean path length				Clustering coefficient			
	beta	SE	t	p	beta	SE	t	p
Intercept	0.599	0.12	5.143	<0.001	0.770	0.02	34.163	<0.001
Real vs. Erdős–Rényi	1.915	0.06	33.152	<0.001	-0.673	0.01	-72.649	<0.001
Real vs. Watts-Strogatz	0.321	0.06	5.668	<0.001	0.156	0.01	17.217	<0.001
Corpus	0.468	0.05	9.813	<0.001	-0.014	0.01	-1.228	0.246
Age	0.000	0.00	-0.027	0.979	0.002	0.00	1.888	0.064
Network size	NA	NA	NA	NA	0.000	0.00	-10.470	<0.001

Table 5

Outputs from linear mixed effects regression models testing comparisons of Actual vs. Target data on mean path length and clustering coefficient.

Effect	Mean path length				Clustering coefficient			
	beta	SE	t	p	beta	SE	t	p
Intercept	-0.080	0.10	-0.801	0.446	0.9340	0.03	32.438	<0.001
Actual vs. Target	0.192	0.01	16.375	<0.001	-0.1143	0.01	-21.043	<0.001
Corpus	1.600	0.04	45.453	<0.001	-0.0274	0.01	-3.790	<0.001
Age	0.005	0.00	1.348	0.195	-0.0056	0.00	-4.453	<0.001
Network size	0.000	0.00	4.233	0.001	-0.0004	0.00	-14.009	<0.001

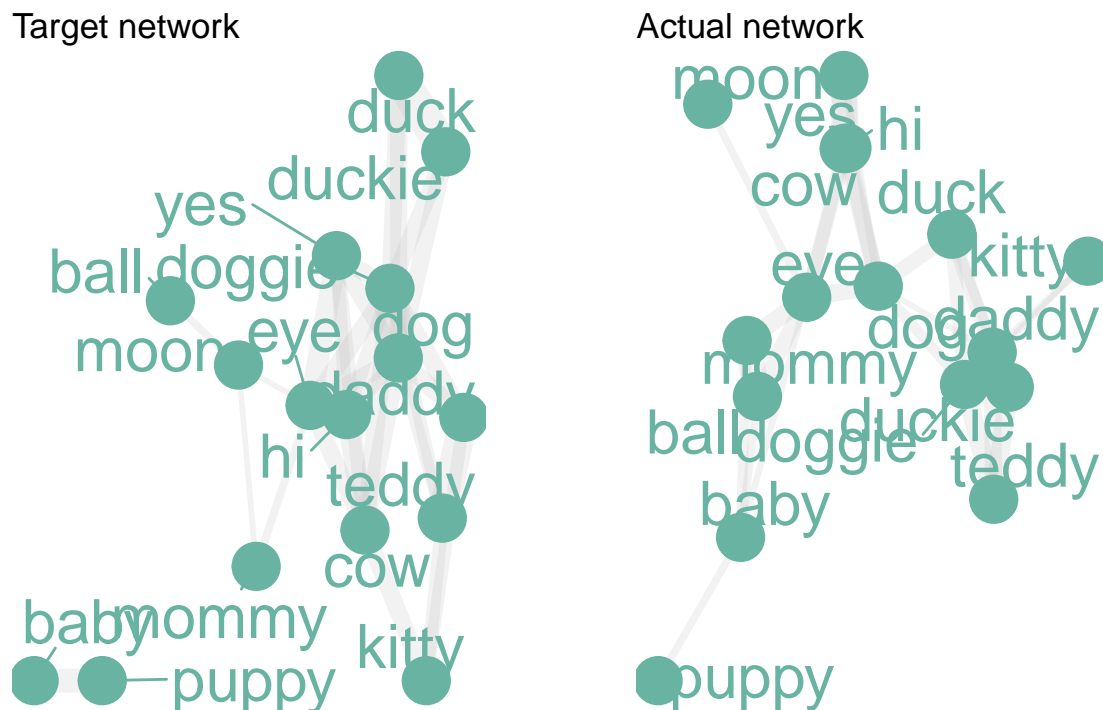
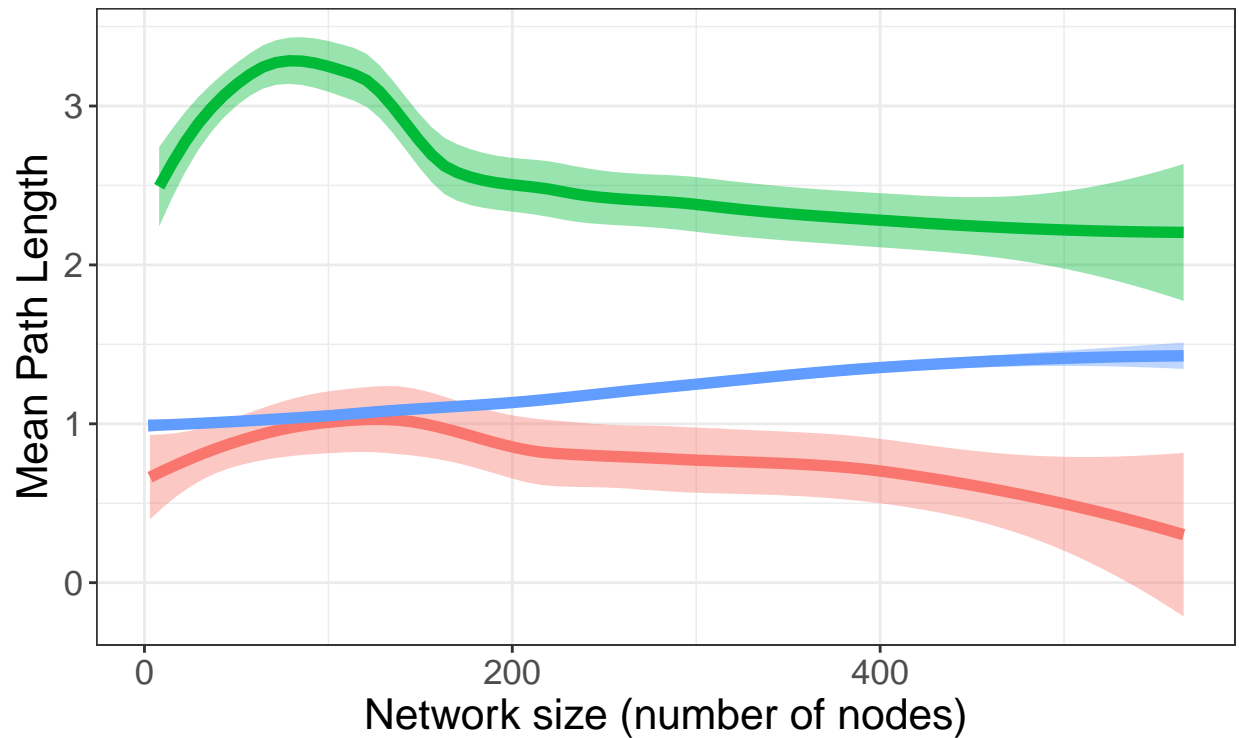
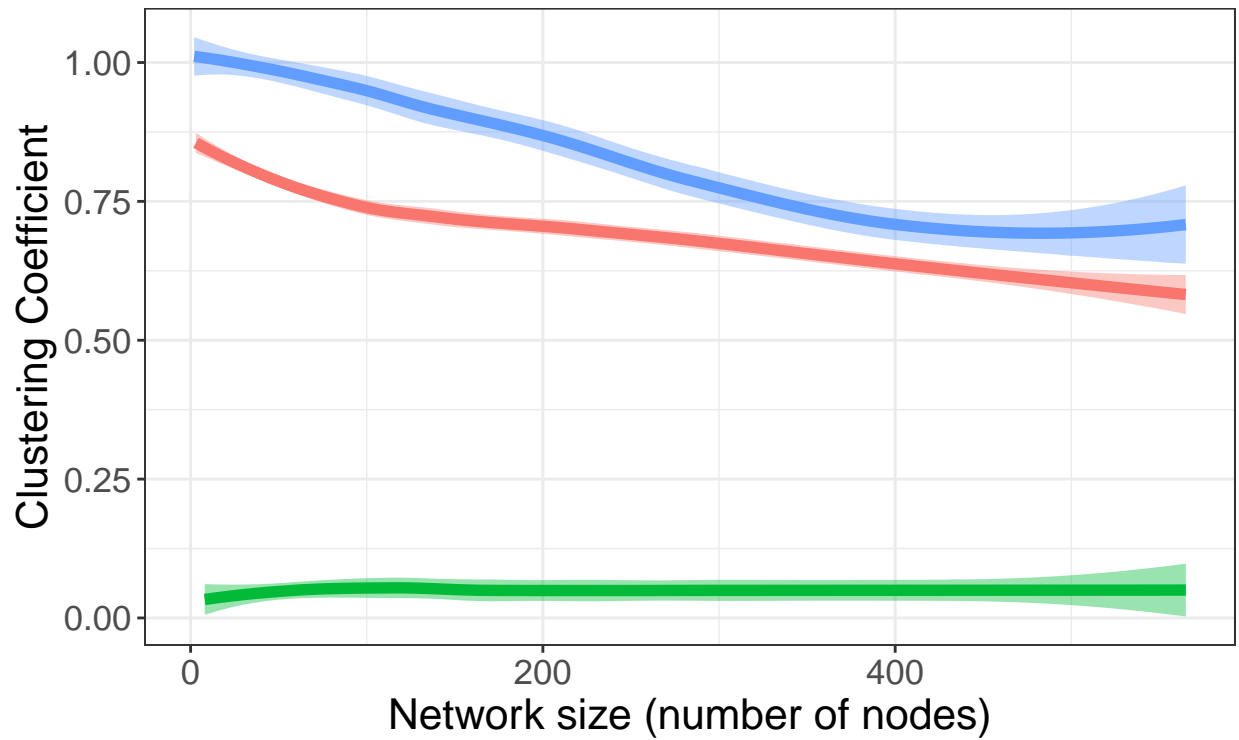


Figure 1. Two networks generated from one English-acquiring infant's data. The networks incorporate all words produced by the child upto and including age 14 months. The Target network is shown on the left, and the Actual network on the right. Teal dots represent nodes (words) in the network, and lines between the dots are the edges. Thicker lines represent shorter phonological distances (these are also visually represented in graph space, whereby closer phonological distance is represented by nodes plotted more closely together). Words represent the target words, though note that the Actual network plots distance between words as produced by the infant. Hermit words had a scaled distance of $>.25$ to all other words in the network.



e ■ Real ■ Simulated – random (Erdos Renyi) ■ Simulated – small world (\

Figure 2. Change in mean path length as network size increases, in Real vs. Simulated (random and small-world) data. Coloured lines represent Data type; coloured bands represent 95% CIs. See S2 for the same data plotted according to the effect of Age.



pe — Real — Simulated – random (Erdos Renyi) — Simulated – small world

Figure 3. Change in clustering coefficient as network size increases, in Real vs. Simulated (random and small-world) data. Coloured lines represent Data type; coloured bands represent 95% CIs. See S2 for the same data plotted according to the effect of Age.

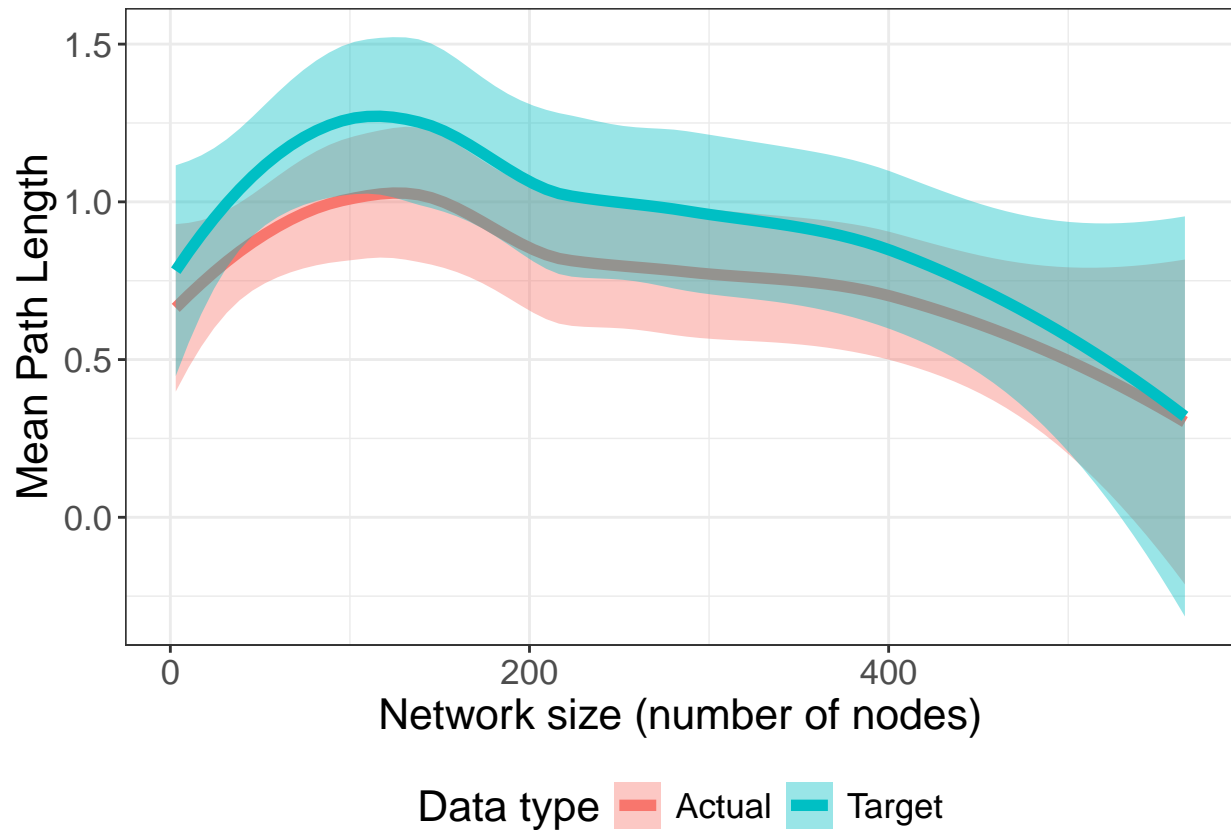


Figure 4. Change in mean path length as network size increases, in Actual vs. Target data. Coloured lines represent Data type; coloured bands represent 95% CIs. See S2 for the same data plotted according to the effect of Age.

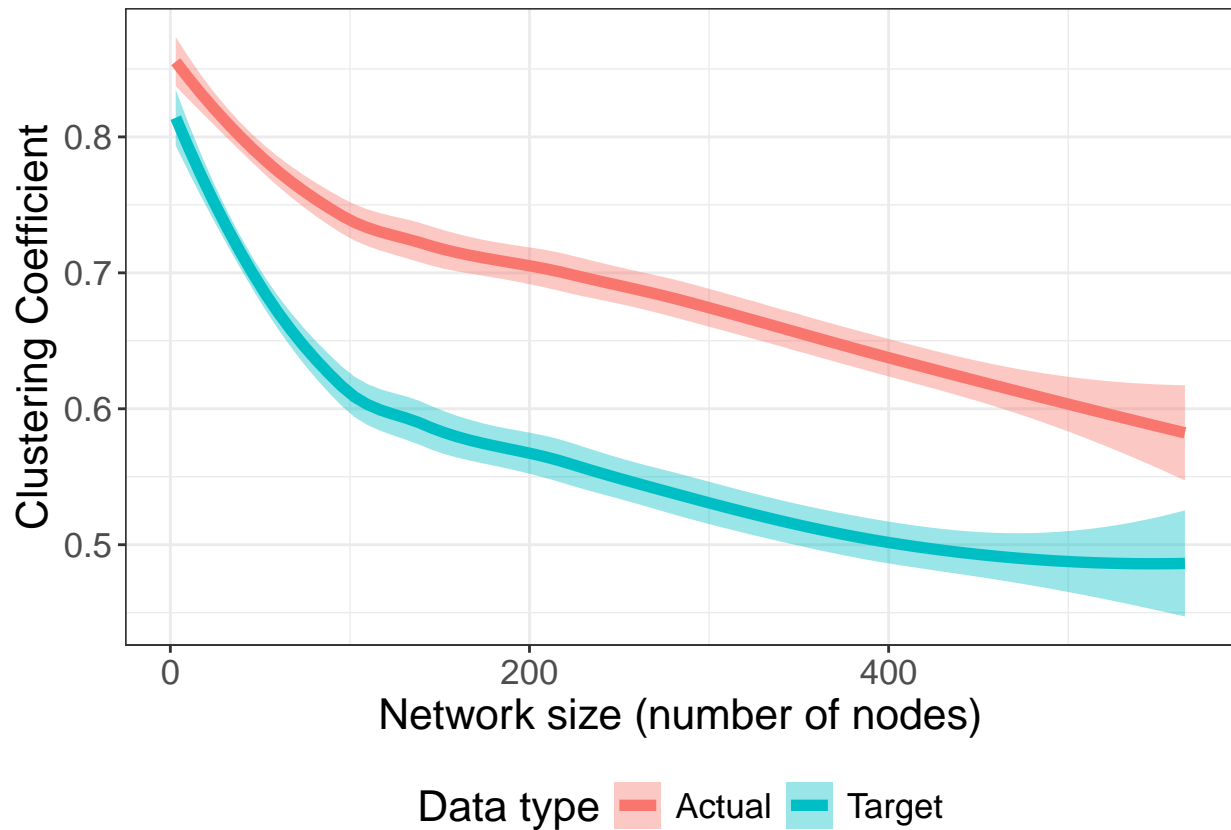


Figure 5. Change in mean clustering coefficient as network size increases, in Actual vs. Target data. Coloured lines represent Data type; coloured bands represent 95% CIs. See S2 for the same data plotted according to the effect of Age.

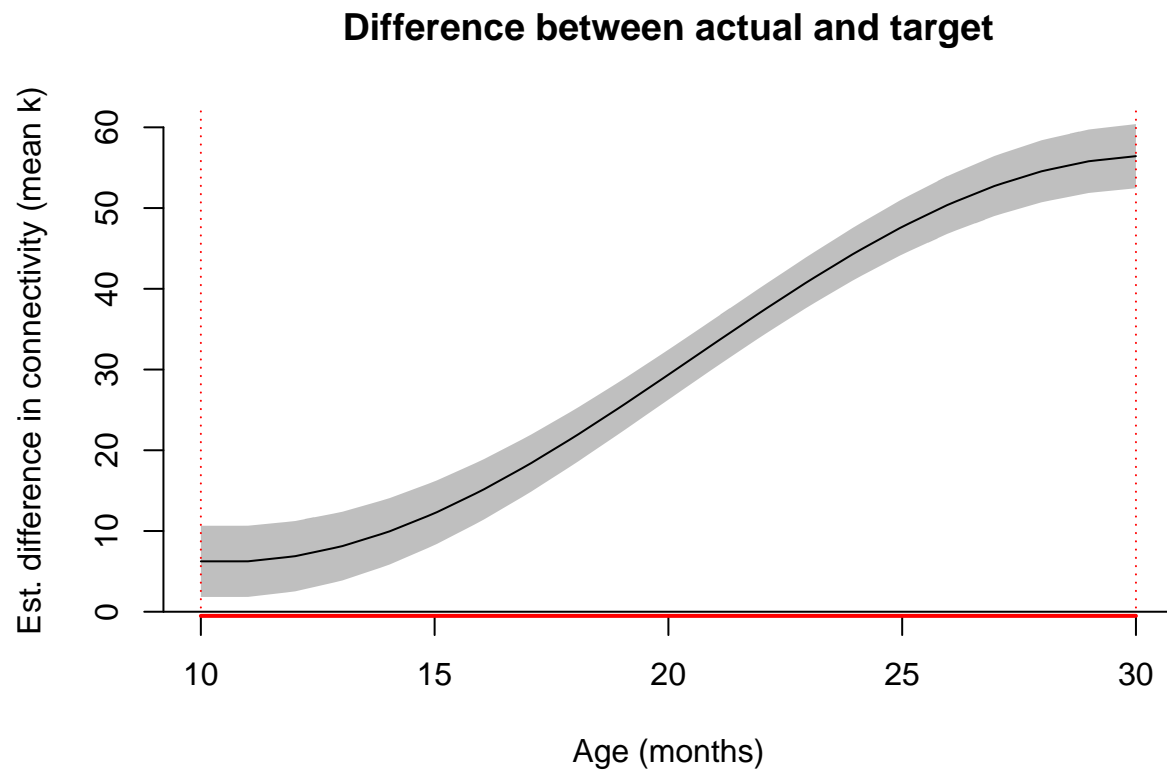


Figure 6. Difference smooth plot showing difference between connectivity (mean k) in Actual vs. Target forms from the GAMM model specified above. Shaded area shows 95% confidence intervals, red line along x-axis indicates months in which the difference between Actual and Target forms was significant.