Modeling the Structure and Dynamics of the Consonant Inventories: A Complex Network Approach

Animesh Mukherjee[1], **Monojit Choudhury**[2], **Anupam Basu**[1], **Niloy Ganguly**[1]
[1]Department of Computer Science and Engineering,
Indian Institute of Technology, Kharagpur, India – 721302
[2]Microsoft Research India, Bangalore, India – 560080
{animeshm,anupam,niloy}@cse.iitkgp.ernet.in,
monojitc@microsoft.com

Abstract

We study the self-organization of the consonant inventories through a complex network approach. We observe that the distribution of occurrence as well as co-occurrence of the consonants across languages follow a power-law behavior. The co-occurrence network of consonants exhibits a high clustering coefficient. We propose four novel synthesis models for these networks (each of which is a refinement of the earlier) so as to successively match with higher accuracy (a) the above mentioned topological properties as well as (b) the linguistic property of *feature economy* exhibited by the consonant inventories. We conclude by arguing that a possible interpretation of this mechanism of network growth is the process of child language acquisition. Such models essentially increase our understanding of the structure of languages that is influenced by their evolutionary dynamics and this, in turn, can be extremely useful for building future NLP applications.

1 Introduction

A large number of regular patterns are observed across the sound inventories of human languages. These regularities are arguably a consequence of the self-organization that is instrumental in the emergence of these inventories (de Boer, 2000). Many attempts have been made by functional phonologists for explaining this self-organizing behavior through certain general principles such as *maximal perceptual contrast* (Liljencrants and Lindblom, 1972), *ease of articulation* (Lindblom and Maddieson, 1988; de Boer, 2000), and *ease of learnability* (de Boer, 2000). In fact, there are a lot of studies that attempt to explain the emergence of the vowel inventories through the application of one or more of the above principles (Liljencrants and Lindblom, 1972; de Boer, 2000). Some studies have also been carried out in the area of linguistics that seek to reason the observed patterns in the consonant inventories (Trubetzkoy, 1939; Lindblom and Maddieson, 1988; Boersma, 1998; Clements, 2008). Nevertheless, most of these works are confined to certain individual principles rather than formulating a general theory describing the emergence of these regular patterns across the consonant inventories.

The self-organization of the consonant inventories emerges due to an interaction of different forces acting upon them. In order to identify the nature of these interactions one has to understand the growth dynamics of these inventories. The theories of *complex networks* provide a number of growth models that have proved to be extremely successful in explaining the evolutionary dynamics of various social (Newman, 2001; Ramasco et al., 2004), biological (Jeong et al., 2000) and other natural systems. The basic framework for the current study develops around two such complex networks namely, the **P**honeme-**La**nguage **Net**work or **PlaNet** (Choudhury et al., 2006) and its one-mode projection, the **Pho**neme-Phoneme **Net**work or **PhoNet** (Mukherjee et al.2007a). We begin by analyzing some of the structural properties (Sec. 2) of the networks and observe that the consonant nodes in both PlaNet and PhoNet follow a power-law-like degree distribution. Moreover, PhoNet is characterized by a high clustering coefficient, a property that has been found to be prevalent in many other social networks (Newman, 2001; Ramasco et al., 2004).

We propose four synthesis models for PlaNet (Sec. 3), each of which employ a variant of a *preferential attachment* (Barabási and Albert, 1999) based growth kernel[1]. While the first two models are independent of the characteristic properties of the (consonant) nodes, the following two use them. These models are successively refined not only to reproduce the topological properties of PlaNet and PhoNet, but also to match the linguistic property of *feature economy* (Boersma, 1998; Clements, 2008) that is observed across the consonant inventories. The underlying growth rules for each of these individual models helps us to interpret the cause of the emergence of at least one (or more) of the aforementioned properties. We conclude (Sec. 4) by providing a possible interpretation of the proposed mathematical model that we finally develop in terms of child language acquisition.

There are three major contributions of this work. Firstly, it provides a fascinating account of the structure and the evolution of the human speech sound systems. Furthermore, the introduction of the node property based synthesis model is a significant contribution to the field of complex networks. On a broader perspective, this work shows how statistical mechanics can be applied in understanding the structure of a linguistic system, which in turn can be extremely useful in developing future NLP applications.

2 Properties of the Consonant Inventories

In this section, we briefly recapitulate the definitions of PlaNet and PhoNet, the data source, construction procedure for the networks and some of their important structural properties. We also revisit the concept of feature economy and the method used for its quantification.

[1]The word kernel here refers to the function or mathematical formula that drives the growth of the network.

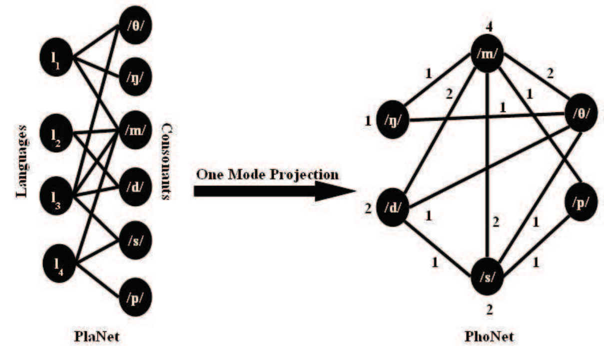

Figure 1: Illustration of the nodes and edges of PlaNet and PhoNet.

2.1 Structural Properties of the Consonant Networks

PlaNet is a bipartite graph $G = \langle V_L, V_C, E_{pl} \rangle$ consisting of two sets of nodes namely, V_L (labeled by the languages) and V_C (labeled by the consonants); E_{pl} is the set of edges running between V_L and V_C. There is an edge $e \in E_{pl}$ from a node $v_l \in V_L$ to a node $v_c \in V_C$ iff the consonant c is present in the inventory of language l.

PhoNet is the one-mode projection of PlaNet onto the consonant nodes i.e., a network of consonants in which two nodes are linked by an edge with weight as many times as they co-occur across languages. Hence, it can be represented by a graph $G = \langle V_C, E_{ph} \rangle$, where V_C is the set of consonant nodes and E_{ph} is the set of edges connecting these nodes in G. There is an edge $e \in E_{ph}$ if the two nodes (read consonants) that are connected by e co-occur in at least one language and the number of languages they co-occur in defines the weight of the edge e. Figure 1 shows the nodes and the edges of PlaNet and PhoNet.

Data Source and Network Construction: Like many other earlier studies (Liljencrants and Lindblom, 1972; Lindblom and Maddieson, 1988; de Boer, 2000; Hinskens and Weijer, 2003), we use the UCLA Phonological Segment Inventory Database (UPSID) (Maddieson, 1984) as the source of our data. There are 317 languages in the database with a total of 541 consonants found across them. Each consonant is characterized by a set of phonological features (Trubetzkoy, 1931), which distinguishes

Manner of Articulation	Place of Articulation	Phonation
tap	velar	voiced
flap	uvular	voiceless
trill	dental	
click	palatal	
nasal	glottal	
plosive	bilabial	
r-sound	alveolar	
fricative	retroflex	
affricate	pharyngeal	
implosive	labial-velar	
approximant	labio-dental	
ejective stop	labial-palatal	
affricated click	dental-palatal	
ejective affricate	dental-alveolar	
ejective fricative	palato-alveolar	
lateral approximant		

Table 1: The table shows some of the important features listed in UPSID. Over 99% of the UPSID languages have bilabial, dental-alveolar and velar plosives. Furthermore, voiceless plosives outnumber the voiced ones (92% vs. 67%). 93% of the languages have at least one fricative, 97% have at least one nasal and 96% have at least one liquid. Approximants occur in fewer than 95% of the languages.

it from others. UPSID uses articulatory features to describe the consonants, which can be broadly categorized into three different types namely the *manner of articulation*, the *place of articulation* and *phonation*. Manner of articulation specifies how the flow of air takes place in the vocal tract during articulation of a consonant, whereas place of articulation specifies the active speech organ and also the place where it acts. Phonation describes the vibration of the vocal cords during the articulation of a consonant. Apart from these three major classes there are also some secondary articulatory features found in certain languages. There are around 52 features listed in UPSID; the important ones are noted in Table 1. Note that in UPSID the features are assumed to be binary-valued and therefore, each consonant can be represented by a binary vector.

We have used UPSID in order to construct PlaNet and PhoNet. Consequently, $|V_L|$ = 317 (in PlaNet) and $|V_C|$ = 541. The number of edges in PlaNet and PhoNet are 7022 and 30412 respectively.

Degree Distributions of PlaNet and PhoNet: The degree distribution is the fraction of nodes, denoted by P_k, which have a degree[2] greater than or equal to k (Newman, 2003). The degree distribution of the consonant nodes in PlaNet and PhoNet are shown in Figure 2 in the log-log scale. Both the plots show a power-law behavior ($P_k \propto k^{-\alpha}$) with exponential cut-offs towards the ends. The value of α is 0.71 for PlaNet and 0.89 for PhoNet.

Clustering Coefficient of PhoNet: The clustering coefficient for a node i is the proportion of links between the nodes that are the neighbors of i divided by the number of links that could possibly exist between them (Newman, 2003). Since PhoNet is a weighted graph the above definition is suitably modified by the one presented in (Barrat et al., 2004). According to this definition, the clustering coefficient for a node i is,

$$c_i = \frac{1}{\left(\sum_{\forall j} w_{ij}\right)(k_i - 1)} \sum_{\forall j,l} \frac{(w_{ij} + w_{il})}{2} a_{ij} a_{il} a_{jl} \tag{1}$$

where j and l are neighbors of i; k_i represents the plain degree of the node i; w_{ij}, w_{jl} and w_{il} denote the weights of the edges connecting nodes i and j, j and l, and i and l respectively; a_{ij}, a_{il}, a_{jl} are boolean variables, which are true iff there is an edge between the nodes i and j, i and l, and j and l respectively. The clustering coefficient of the network (c_{av}) is equal to the average clustering coefficient of the nodes. The value of c_{av} for PhoNet is 0.89, which is significantly higher than that of a random graph with the same number of nodes and edges (0.08).

2.2 Linguistic Properties: Feature Economy and its Quantification

The principle of feature economy states that languages tend to use a small number of *distinctive* features and maximize their combinatorial possibilities to generate a large number of consonants (Boersma, 1998; Clements, 2008). Stated differently, a given consonant will have a higher than expected chance of occurrence in inventories in which all of its features have already distinctively occurred in the other consonants. This principle immediately implies that the consonants chosen by a language should share a considerable number of features among them. The quantification process, which is a refinement of the idea presented in (Mukherjee et al.2007b), is as follows.

Feature Entropy: For an inventory of size N, let there be p_f consonants for which a particular feature

[2]For a weighted graph like PhoNet, the degree of a node i is the sum of weights on the edges that are incident on i.

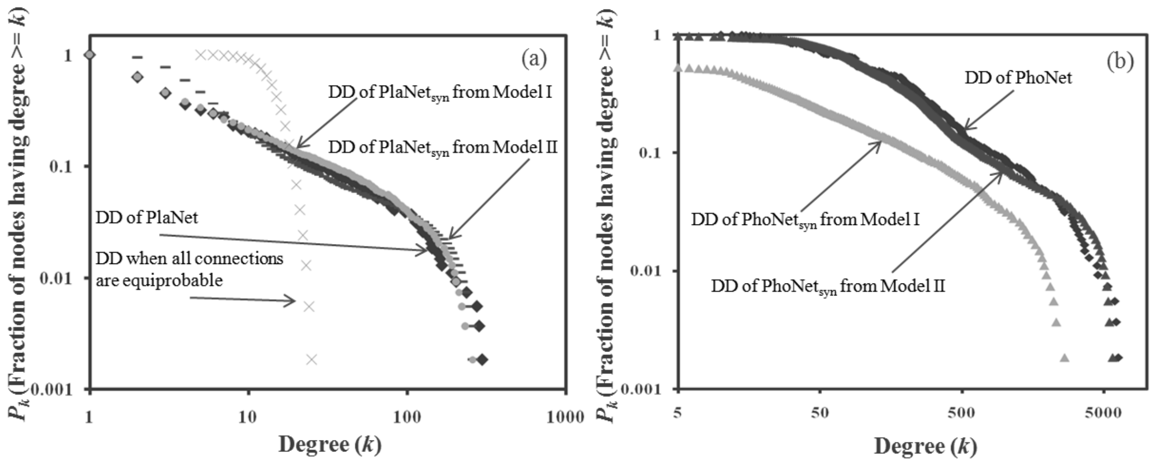

Figure 2: Degree distribution (DD) of PlaNet along with that of PlaNet$_{syn}$ obtained from Model I and II respectively; (b) DD of PhoNet along with that of PhoNet$_{syn}$ obtained from Model I and II respectively. Both the plots are in log-log scale.

f (recall that we assume f to be binary-valued) is present and q_f other consonants for which the same is absent. Therefore, the probability that a consonant (chosen uniformly at random from this inventory) contains the feature f is $\frac{p_f}{N}$ and the probability that it does not contain the feature is $\frac{q_f}{N}$ (=1–$\frac{p_f}{N}$). One can think of f as an independent random variable, which can take values 1 and 0, and $\frac{p_f}{N}$ and $\frac{q_f}{N}$ define the probability distribution of f. Therefore, for any given inventory, we can define the binary entropy H_f (Shannon and Weaver, 1949) for the feature f as

$$H_f = -\frac{p_f}{N}\log_2\frac{p_f}{N} - \frac{q_f}{N}\log_2\frac{q_f}{N} \qquad (2)$$

If F is the set of all features present in the consonants forming the inventory, then *feature entropy* F_E is the sum of the binary entropies with respect to all the features, that is

$$F_E = \sum_{f\in F} H_f = \sum_{f\in F}\left(-\frac{p_f}{N}\log_2\frac{p_f}{N} - \frac{q_f}{N}\log_2\frac{q_f}{N}\right) \qquad (3)$$

Since we have assumed that f is an independent random variable, F_E is the joint entropy of the system. In other words, F_E provides an estimate of the number of discriminative features present in the consonants of an inventory that a speaker (e.g., parent) has to communicate to a learner (e.g., child) during language transmission. The lower the value of F_E the higher is the feature economy. The curve marked as (R) in Figure 3 shows the average feature entropy of the consonant inventories of a particular size[3] (y-axis) versus the inventory size (x-axis).

3 Synthesis Models

In this section, we describe four synthesis models that incrementally attempt to explain the emergence of the structural properties of PlaNet and PhoNet as well as the feature entropy exhibited by the consonant inventories. In all these models, we assume that the distribution of the consonant inventory size, i.e., the degrees of the language nodes in PlaNet, are known *a priori*.

3.1 Model I: Preferential Attachment Kernel

This model employs a modified version of the kernel described in (Choudhury et al., 2006), which is the only work in literature that attempts to explain the emergence of the consonant inventories in the framework of complex networks.

Let us assume that a language node $L_i \in V_L$ has a degree k_i. The consonant nodes in V_C are assumed to be unlabeled, i.e, they are not marked by the distinctive features that characterize them. We first sort

[3]Let there be n inventories of a particular size k. The average feature entropy of the inventories of size k is $\frac{1}{n}\sum_{i=1}^{n} F_{E_i}$, where F_{E_i} signifies the feature entropy of the i^{th} inventory of size k.

the nodes L_1 through L_{317} in the ascending order of their degrees. At each time step a node L_j, chosen in order, preferentially attaches itself with k_j *distinct* nodes (call each such node C_i) of the set V_C. The probability $Pr(C_i)$ with which the node L_j attaches itself to the node C_i is given by,

$$Pr(C_i) = \frac{{d_i}^{\alpha} + \epsilon}{\sum_{i' \in V_C'} ({d_{i'}}^{\alpha} + \epsilon)} \tag{4}$$

where, d_i is the current degree of the node C_i, $V_C^{'}$ is the set of nodes in V_C that are not already connected to L_j, ϵ is the smoothing parameter that facilitates random attachments and α indicates whether the attachment kernel is sub-linear ($\alpha < 1$), linear ($\alpha = 1$) or super-linear ($\alpha > 1$). Note that the modification from the earlier kernel (Choudhury et al., 2006) is brought about by the introduction of α. The above process is repeated until all the language nodes $L_j \in V_L$ get connected to k_j consonant nodes (refer to Figure. 6 of (Choudhury et al., 2006) for an illustration of the steps of the synthesis process). Thus, we have the synthesized version of PlaNet, which we shall call PlaNet$_{syn}$ henceforth.

The Simulation Results: We simulate the above model to obtain PlaNet$_{syn}$ for 100 different runs and average the results over all of them. We find that the degree distributions that emerge fit the empirical data well for $\alpha \in$ [1.4,1.5] and $\epsilon \in$ [0.4,0.6], the best being at α = 1.44 and ϵ = 0.5 (shown in Figure 2). In fact, the mean error[4] between the real and the synthesized distributions for the best choice of parameters is as small as 0.01. Note that this error in case of the model presented in (Choudhury et al., 2006) was 0.03. Furthermore, as we shall see shortly, a superlinear kernel can explain various other topological properties more accurately than a linear kernel.

In absence of preferential attachment i.e., when all the connections to the consonant nodes are equiprobable, the mean error rises to 0.35.

A possible reason behind the success of this model is the fact that language is a constantly changing system and preferential attachment plays a significant role in this change. For instance, during the change those consonants that belong to languages that are more prevalent among the speakers of a generation have higher chances of being transmitted to the speakers of the subsequent generations (Blevins, 2004). This heterogeneity in the choice of the consonants manifests itself as preferential attachment. We conjecture that the value of α is a function of the societal structure and the cognitive capabilities of human beings. The exact nature of this function is currently not known and a topic for future research. The parameter ϵ in this case may be thought of as modeling the randomness of the system.

Nevertheless, the degree distribution of PhoNet$_{syn}$, which is the one-mode projection of PlaNet$_{syn}$, does not match the real data well (see Figure 2). The mean error between the two distributions is 0.45. Furthermore, the clustering coefficient of PhoNet$_{syn}$ is 0.55 and differs largely from that of PhoNet. The primary reason for this deviation in the results is that PhoNet exhibits strong patterns of co-occurrences (Mukherjee et al.2007a) and this fact is not taken into account by Model I. In order to circumvent the above problem, we introduce the concept of *triad* (i.e., fully connected triplet) formation and thereby refine the model in the following section.

3.2 Model II: Kernel based on Triad Formation

The triad model (Peltomäki and Alava, 2006) builds up on the concept of *neighborhood* formation. Two consonant nodes C_1 and C_2 become neighbors if a language node at any step of the synthesis process attaches itself to both C_1 and C_2. Let the probability of triad formation be denoted by p_t. At each time step a language node L_j (chosen from the set of language nodes sorted in ascending order of their degrees) makes the first connection preferentially to a consonant node $C_i \in V_C$ to which L_j is not already connected following the distribution $Pr(C_i)$. For the rest of the (k_j-1) connections L_j attaches itself preferentially to only the neighbors of C_i to which L_j is not yet connected with a probability p_t. Consequently, L_j connects itself preferentially to the non-neighbors of C_i to which L_j is not yet connected with a probability $(1 - p_t)$. The neighbor set of C_i gets updated accordingly. Note that every time the node C_i and its neighbors are chosen they together

[4] Mean error is defined as the average difference between the ordinate pairs (say y and $y^{'}$) where the abscissas are equal. In other words, if there are N such ordinate pairs then the mean error can be expressed as $\frac{\sum |y-y^{'}|}{N}$.

impose a clique on the one-mode projection. This phenomenon leads to the formation of a large number of triangles in the one-mode projection thereby increasing the clustering coefficient of the resultant network.

The Simulation Results: We carry out 100 different simulation runs of the above model for a particular set of parameter values to obtain PlaNet$_{syn}$ and average the results over all of them. We explore several parameter settings in the range as follows: $\alpha \in$ [1,1.5] (in steps of 0.1), $\epsilon \in$ [0.2,0.4] (in steps of 0.1) and $p_t \in$ [0.70,0.95] (in steps of 0.05). We also observe that if we traverse any further along one or more of the dimensions of the parameter space then the results get worse. The best result emerges for α = 1.3, ϵ = 0.3 and p_t = 0.8.

Figure 2 shows the degree distribution of the consonant nodes of PlaNet$_{syn}$ and PlaNet. The mean error between the two distributions is 0.04 approximately and is therefore worse than the result obtained from Model I. Nevertheless, the average clustering coefficient of PhoNet$_{syn}$ in this case is 0.85, which is within 4.5% of that of PhoNet. Moreover, in this process the mean error between the degree distribution of PhoNet$_{syn}$ and PhoNet (as illustrated in Figure 2) has got reduced drastically from 0.45 to 0.03.

One can again find a possible association of this model with the phenomena of language change. If a group of consonants largely co-occur in the languages of a generation of speakers then it is very likely that all of them get transmitted together in the subsequent generations (Blevins, 2004). The triad formation probability ensures that if a pair of consonant nodes become neighbors of each other in a particular step of the synthesis process then the choice of such a pair should be highly preferred in the subsequent steps of the process. This is coherent with the aforementioned phenomenon of transmission of consonants in groups over linguistic generations. Since the value of p_t that we obtain is quite high, it may be argued that such transmissions are largely prevalent in nature.

Although Model II reproduces the structural properties of PlaNet and PhoNet quite accurately, as we shall see shortly, it fails to generate inventories that closely match the real ones in terms of feature entropy. However, at this point, recall that Model II assumes that the consonant nodes are unlabeled; therefore, the inventories that are produced as a result of the synthesis are composed of consonants, which unlike the real inventories, are not marked by their distinctive features. In order to label them we perform the following,

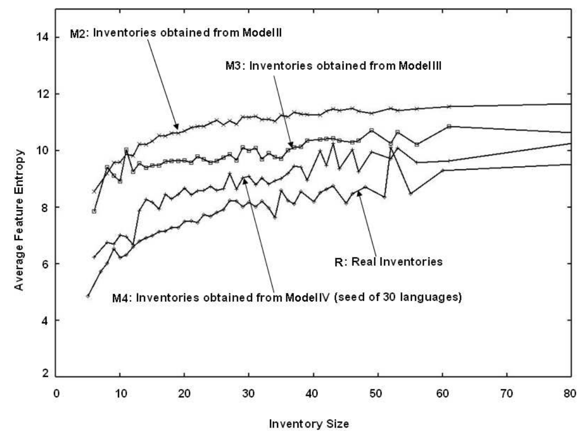

Figure 3: Average feature entropy of the inventories of a particular size (y-axis) versus the inventory size (x-axis).

The Labeling Scheme:

1. Sort the consonants of UPSID in the decreasing order of their frequency of occurrence and call this list of consonants $ListC[1 \cdots 541]$,

2. Sort the V_C nodes of PlaNet$_{syn}$ in decreasing order of their degree and call this list of nodes $ListN[1 \cdots 541]$,

3. $\forall_{1 \leq i \leq 541}$ $ListN[i] \longleftarrow ListC[i]$

The Figure 3 indicates that the curve for the real inventories (R) and those obtained from Model II (M2) are significantly different from each other. This difference arises due to the fact that in Model II, the choice of a consonant from the set of neighbors is solely degree-dependent, where the relationships between the features are not taken into consideration. Therefore, in order to eliminate this problem, we introduce the model using the feature-based kernel in the next section.

3.3 Model III: Feature-based Kernel

In this model, we assume that each of the consonant nodes are labeled, that is each of them are marked by a set of distinctive features. The attachment kernel in this case has two components one of which is preferential while the other favors the choice of those consonants that are at a low feature distance (the number of feature positions they differ at) from the already chosen ones. Let us denote the feature distance between two consonants C_i and $C_i^{'}$ by $D(C_i, C_i^{'})$. We define the *affinity*, $A(C_i, C_i^{'})$, between C_i and $C_i^{'}$ as

$$A(C_i, C_i^{'}) = \frac{1}{D(C_i, C_i^{'})} \quad (5)$$

Therefore, the lower the feature distance between C_i and $C_i^{'}$ the higher is the affinity between them.

At each time step a language node establishes the first connection with a consonant node (say C_i) preferentially following the distribution $Pr(C_i)$ like the previous models. The rest of the connections to any arbitrary consonant node $C_i^{'}$ (not yet connected to the language node) are made following the distribution $(1-w)Pr(C_i^{'}) + wPr_{aff}(C_i, C_i^{'})$, where

$$Pr_{aff}(C_i, C_i^{'}) = \frac{A(C_i, C_i^{'})}{\sum_{\forall C_i^{'}} A(C_i, C_i^{'})} \quad (6)$$

and $0 < w < 1$.

Simulation Results: We perform 100 different simulation runs of the above model for a particular set of parameter values to obtain PlaNet$_{syn}$ and average the results over all of them. We explore different parameter settings in the range as follows: $\alpha \in$ [1,2] (in steps of 0.1), $\epsilon \in$ [0.1,1] (in steps of 0.1) and $w \in$ [0.1,0.5] (in steps of 0.05). The best result in terms of the structural properties of PlaNet and PhoNet emerges for α = 1.6, ϵ = 0.3 and w = 0.2.

In this case, the mean error between the degree distribution curves for PlaNet$_{syn}$ and PlaNet is 0.05 and that between of PhoNet$_{syn}$ and PhoNet is 0.02. Furthermore, the clustering coefficient of PhoNet$_{syn}$ in this case is 0.84, which is within 5.6% of that of PhoNet. The above results show that the structural properties of the synthesized networks in this case are quite similar to those obtained through the triad model. Nevertheless, the average feature entropy of the inventories produced (see curve M3 in Figure 3) are more close to that of the real ones now (for quantitative comparison see Table 2).

Therefore, it turns out that the groups of consonants that largely co-occur in the languages of a linguistic generation are actually driven by the principle of feature economy (see (Clements, 2008; Mukherjee et al.2007a) for details).

Results	Model I	Model II	Model III	Model IV
ME: DD of PlaNet & PlaNet$_{syn}$	0.01	0.04	0.05	0.05
ME: DD of PhoNet & PhoNet$_{syn}$	0.45	0.03	0.02	0.02
% Err: Clustering Coefficient	38.2	04.5	05.6	06.7
ME: Avg. F$_E$ of Real & Synth. Inv.	3.40	3.00	2.10	0.93
α	1.44	1.30	1.60	1.35
ϵ	0.5	0.3	0.3	0.3
p_t	–	0.8	–	–
w	–	–	0.20	0.15

Table 2: Important results obtained from each of the models. ME: Mean Error, DD: Degree Distribution.

However, note that even for Model III the nodes that are chosen for attachment in the initial stages of the synthesis process are arbitrary and consequently, the labels of the nodes of PlaNet$_{syn}$ do not have a one-to-one correspondence with that of PlaNet, which is the main reason behind the difference in the result between them. In order to overcome this problem we can make use of a small set of real inventories to bootstrap the model.

3.4 Model IV: Feature-based Kernel and Bootstrapping

In order to create a bias towards the labeling scheme prevalent in PlaNet, we use 30 (around 10% of the) real languages as a seed (chosen randomly) for Model III; i.e., they are used by the model for bootstrapping. The idea is summarized below.

1. Select 30 real inventories at random and construct a PlaNet from them. Call this network the initial PlaNet$_{syn}$.
2. The rest of the language nodes are incrementally added to this initial PlaNet$_{syn}$ using Model III.

Simulation Results: The best fit now emerges at $\alpha = 1.35$, $\epsilon = 0.3$ and $w = 0.15$. The mean error between the degree distribution of PlaNet and PlaNet$_{syn}$ is 0.05 and that between PhoNet and PhoNet$_{syn}$ is 0.02. The clustering coefficient of PhoNet$_{syn}$ is 0.83 in this case (within 6.7% of that of PhoNet).

The inventories that are produced as a result of the bootstrapping have an average feature entropy closer to the real inventories (see curve M4 in Figure 3) than the earlier models. Hence, we find that this improved labeling strategy brings about a global betterment in our results unlike in the previous cases. The larger the number of languages used for the purpose of bootstrapping the better are the results mainly in terms of the match in the feature entropy curves.

4 Conclusion

We dedicated the preceding sections of this article to analyze and synthesize the consonant inventories of the world's languages in the framework of a complex network. Table 2 summarizes the results obtained from the four models so that the reader can easily compare them. Some of our important observations are

• The distribution of occurrence and co-occurrence of consonants across languages roughly follow a power law,

• The co-occurrence network of consonants has a large clustering coefficient,

• Groups of consonants that largely co-occur across languages are driven by feature economy (which can be expressed through feature entropy),

• Each of the above properties emerges due to different reasons, which are successively unfurled by our models.

So far, we have tried to explain the physical significance of our models in terms of the process of language change. Language change is a collective phenomenon that functions at the level of a population of speakers (Steels, 2000). Nevertheless, it is also possible to explain the significance of the models at the level of an individual, primarily in terms of the process of language acquisition, which largely governs the course of language change. In the initial years of language development every child passes through a stage called *babbling* during which he/she learns to produce non-meaningful sequences of consonants and vowels, some of which are not even used in the language to which they are exposed (Jakobson, 1968; Locke, 1983). Clear preferences can be observed for learning certain sounds such as plosives and nasals, whereas fricatives and liquids are avoided. In fact, this hierarchy of preference during the babbling stage follows the cross-linguistic frequency distribution of the consonants. This innate frequency dependent preference towards certain phonemes might be because of phonetic reasons (i.e., for articulatory/perceptual benefits). In all our models, this innate preference gets captured through the process of preferential attachment. However, at the same time, in the context of learning a particular inventory the ease of learning the individual consonants also plays an important role. The lower the number of new feature distinctions to be learnt, the higher the ease of learning the consonant. Therefore, there are two orthogonal preferences: (a) the occurrence frequency dependent preference (that is innate), and (b) the feature-dependent preference (that increases the ease of learning), which are instrumental in the acquisition of the inventories. The feature-based kernel is essentially a linear combination of these two mutually orthogonal factors.

References

[Barabási and Albert1999] A.-L. Barabási and R. Albert. 1999. Emergence of scaling in random networks. *Science* **286**, 509-512.

[Barrat et al.2004] A. Barrat, M. Barthélemy, R. Pastor-Satorras and A. Vespignani. 2004. The architecture of complex weighted networks. *PNAS* **101**, 3747–3752.

[Blevins2004] J. Blevins. 2004. *Evolutionary Phonology: The Emergence of Sound Patterns*, Cambridge University Press, Cambridge.

[de Boer2000] B. de Boer. 2000. Self-organisation in vowel systems. *Journal of Phonetics* **28**(4), 441–465.

[Boersma1998] P. Boersma. 1998. *Functional Phonology*, The Hague: Holland Academic Graphics.

[Choudhury et al.2006] M. Choudhury, A. Mukherjee, A. Basu and N. Ganguly. 2006. Analysis and synthesis of the distribution of consonants over languages: A complex network approach. *Proceedings of COLING-ACL06*, 128–135.

[Clements2008] G. N. Clements. 2008. The role of features in speech sound inventories. In Eric Raimy & Charles Cairns, eds.,*Contemporary Views on Architecture and Representations in Phonological Theory*, Cambridge, MA: MIT Press.

[Hinskens and Weijer2003] F. Hinskens and J. Weijer. 2003. Patterns of segmental modification in consonant inventories: A cross-linguistic study. *Linguistics* **41**(6), 1041–1084.

[Jakobson1968] R. Jakobson. 1968. *Child Language, Aphasia and Phonological Universals*. The Hague: Mouton.

[Jeong et al.2000] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A. L. Barabási. 2000. The large-scale organization of metabolic networks. *Nature* **406** 651-654.

[Liljencrants and Lindblom1972] J. Liljencrants and B. Lindblom. 1972. Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language* **48**, 839–862.

[Lindblom and Maddieson1988] B. Lindblom and I. Maddieson. 1988. Phonetic universals in consonant systems. *Language, Speech, and Mind*, 62–78, Routledge, London.

[Locke1983] J. L. Locke. 1983. *Phonological Acquisition and Change*. Academic Press New York.

[Maddieson1984] I. Maddieson. 1984. *Patterns of Sounds*, Cambridge University Press, Cambridge.

[Mukherjee et al.2007a] A. Mukherjee, M. Choudhury, A. Basu and N. Ganguly. 2007a. Modeling the co-occurrence principles of the consonant inventories: A complex network approach. *Int. Jour. of Mod. Phys. C* **18**(2), 281–295.

[Mukherjee et al.2007b] A. Mukherjee, M. Choudhury, A. Basu and N. Ganguly. 2007b. Redundancy ratio: An invariant property of the consonant inventories of the world's languages *Proceedings of ACL07*, 104–111.

[Newman2001] M. E. J. Newman. 2001. Scientific collaboration networks. *Physical Review E* **64**, 016131.

[Newman2003] M. E. J. Newman. 2003. The structure and function of complex networks. *SIAM Review* **45**, 167–256.

[Peltomäki and Alava2006] M. Peltomäki and M. Alava. 2006. Correlations in bipartite collaboration networks. *Journal of Statistical Mechanics: Theory and Experiment*, P01010.

[Ramasco et al.2004] J. J. Ramasco, S. N. Dorogovtsev and R. Pastor-Satorras. 2004. Self-organization of collaboration networks. *Physical Review E* **70**, 036106.

[Shannon and Weaver1949] C. E. Shannon and W. Weaver. 1949. *The Mathematical Theory of Information*. University of Illinois Press, Urbana.

[Steels2000] L. Steels. 2000. Language as a complex adaptive system. In: Schoenauer, M., editor, *Proceedings of PPSN VI*, *LNCS*, 17–26.

[Trubetzkoy1931] N. Trubetzkoy. 1931. Die phonologischen systeme. *TCLP* **4**, 96–116.

[Trubetzkoy1939] N. Trubetzkoy. 1969. *Principles of Phonology*. English translation of *Grundzüge der Phonologie, 1939*, University of California Press, Berkeley.