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

Animesh Mukherjee

Department of Computer Science and Engineering, IIT Kharagpur

Co-authors:

Monojit Choudhury, Microsoft Research India

Anupam Basu, CSE, IIT Kharagpur, India

Niloy Ganguly, CSE, IIT Kharagpur, India

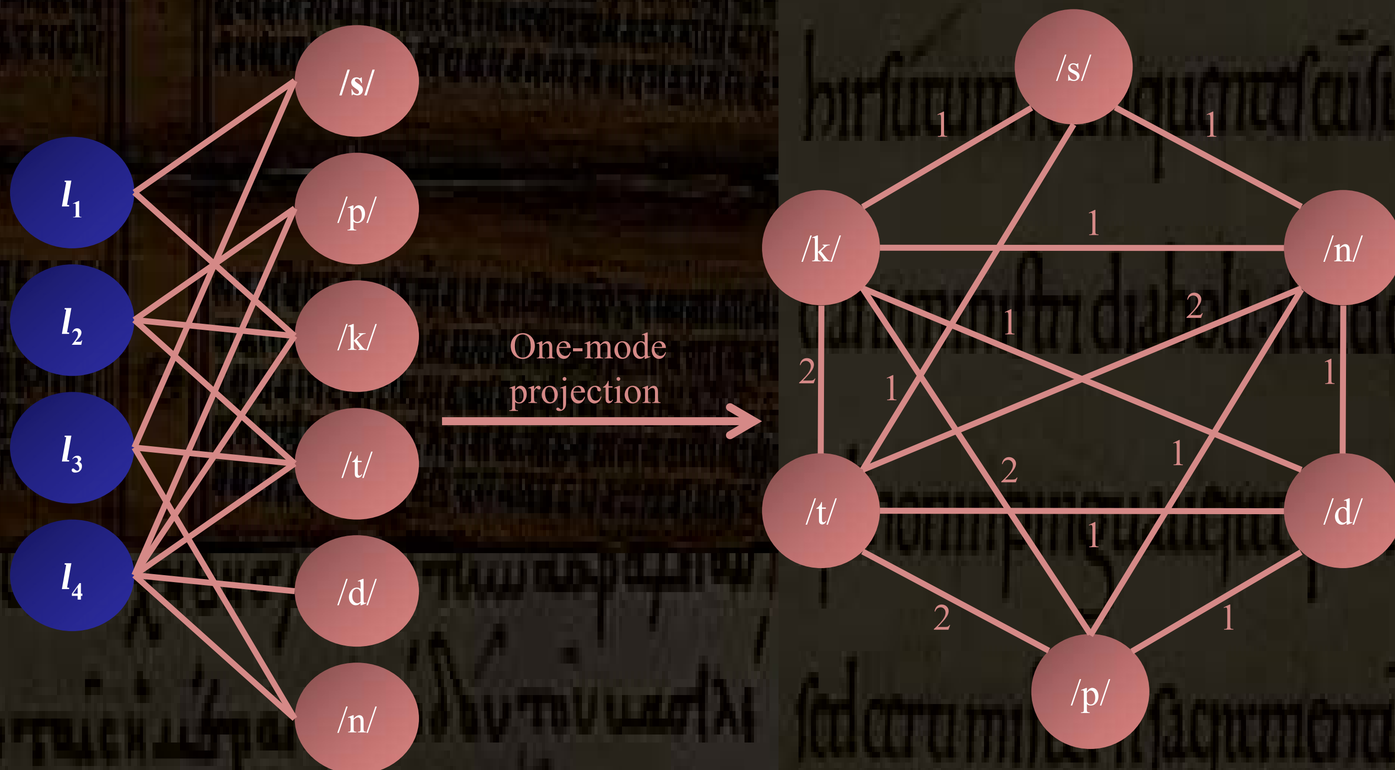
Patterns of Sounds

- Speech sounds are often represented as abstract units called phonemes that are in turn characterized by a set of features
- Sound Inventory – Repertoire of sounds/phonemes that the speakers of a language use to communicate
- Three basic types of inventories: vowels, consonants, and dipthongs
- These inventories show remarkably regular cross-linguistic patterns
- For instance, if a sound inventory has three vowels then in 95% of the cases they are /i/, /a/ and /u/ → for perceptual benefits

Patterns in Consonant Inventories

- A lot of attempts right from 1930s
- Several individual principles proposed but no holistic theory emerged mainly due to the inherent complexity of the problem
  - Larger inventory size
  - Larger set of features
  - Multiple forces acting together to shape them
- *A versatile modeling methodology for extracting and explaining these patterns is hitherto absent*
  → The primary motivation for this study

Network of consonants

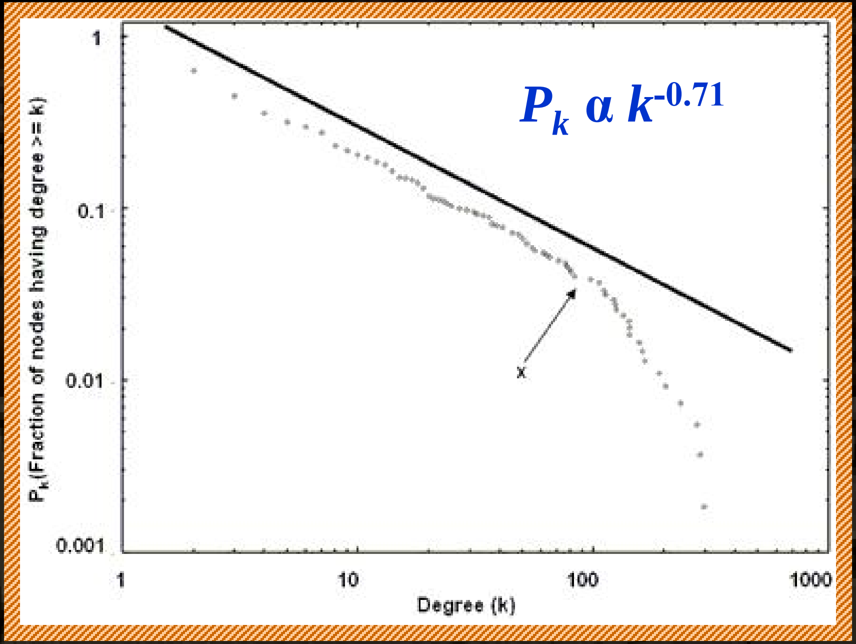


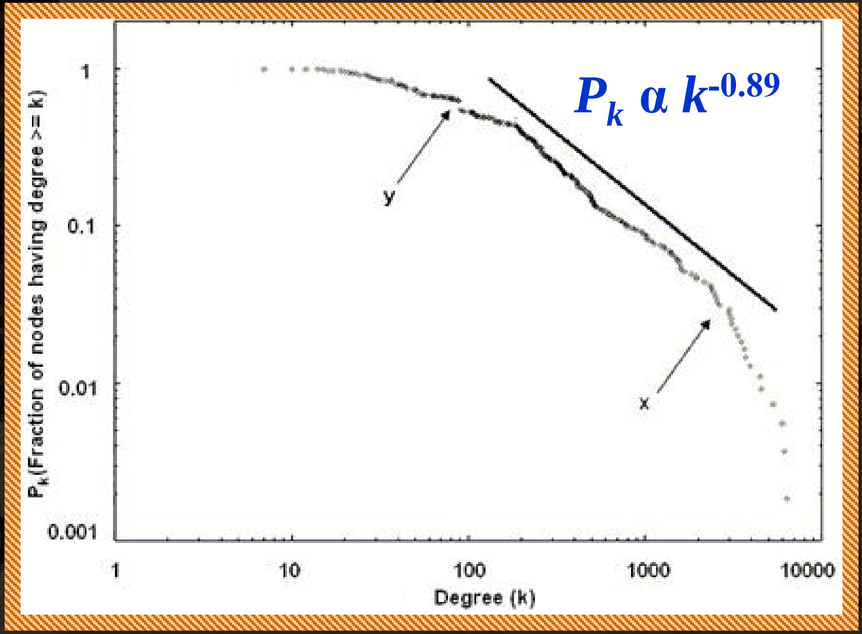
PlaNet - Phoneme-Language Network   PhoNet - Phoneme-Phoneme Network

- Networks constructed from the data available at UCLA Phonological Segment Inventory Database (UPSID) → hosts 317 inventories with 541 different consonants found across them

Structural Properties of the Networks

- Degree Distribution: Fraction of nodes P_k with degree $\geq k$





- Clustering Coefficient: Probability that two of my friends are also friends → many triangles in the network
  - Clustering Coefficient of PhoNet is as high as **0.89** → Many triangles [(/p/,/t/,/k/), (/b/,/d/,/g/) etc.] (see Mukherjee et al. 2007 for reference)

Linguistic Property: Feature Economy

- Consonants tend to occur in pairs that are highly correlated in terms of their features
- Languages tend to maximize combinatorial possibilities of a few features to produce many consonants

plosive	voiced	voiceless
bilabial	/b/	/p/
dental	/d/	/t/

If a language has in its inventory

then it will also tend to have

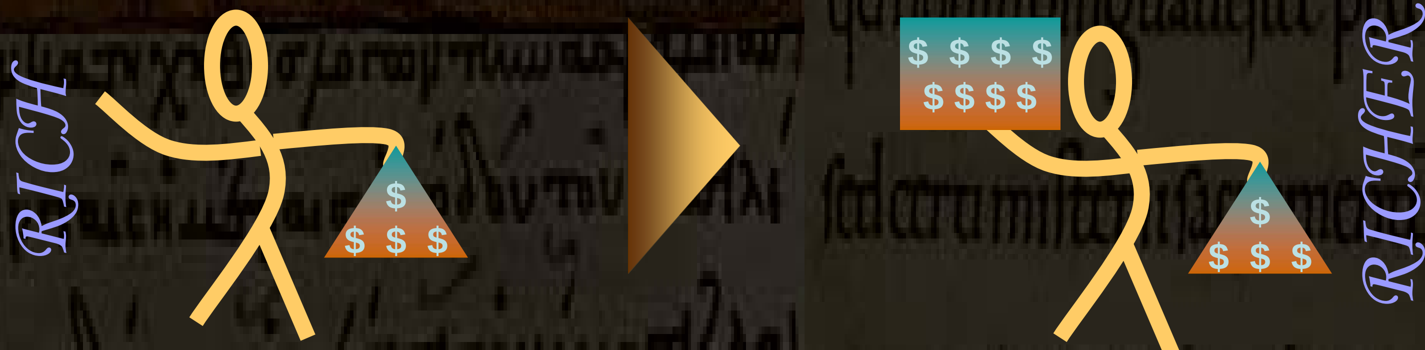
Feature Entropy

- p_f – number of consonants in the inventory in which feature *f* is present
- q_f – number of consonants in the inventory in which feature *f* is absent
- The probability that a consonant chosen at random form the inventory has *f* is $\frac{p_f}{N}$ and that is does not have *f* is $\frac{q_f}{N}$ $(1-\frac{p_f}{N})$
- If F denote the set of all features,

$$F_E = -\sum_{f \epsilon F} \frac{p_f}{N} log_2 \frac{p_f}{N} + \frac{q_f}{N} log_2 \frac{q_f}{N}$$

How do these properties emerge?

- One way to answer this question is to hit upon a model based on network growth
- *Preferential attachment* based models are the most popular "power-law generators"

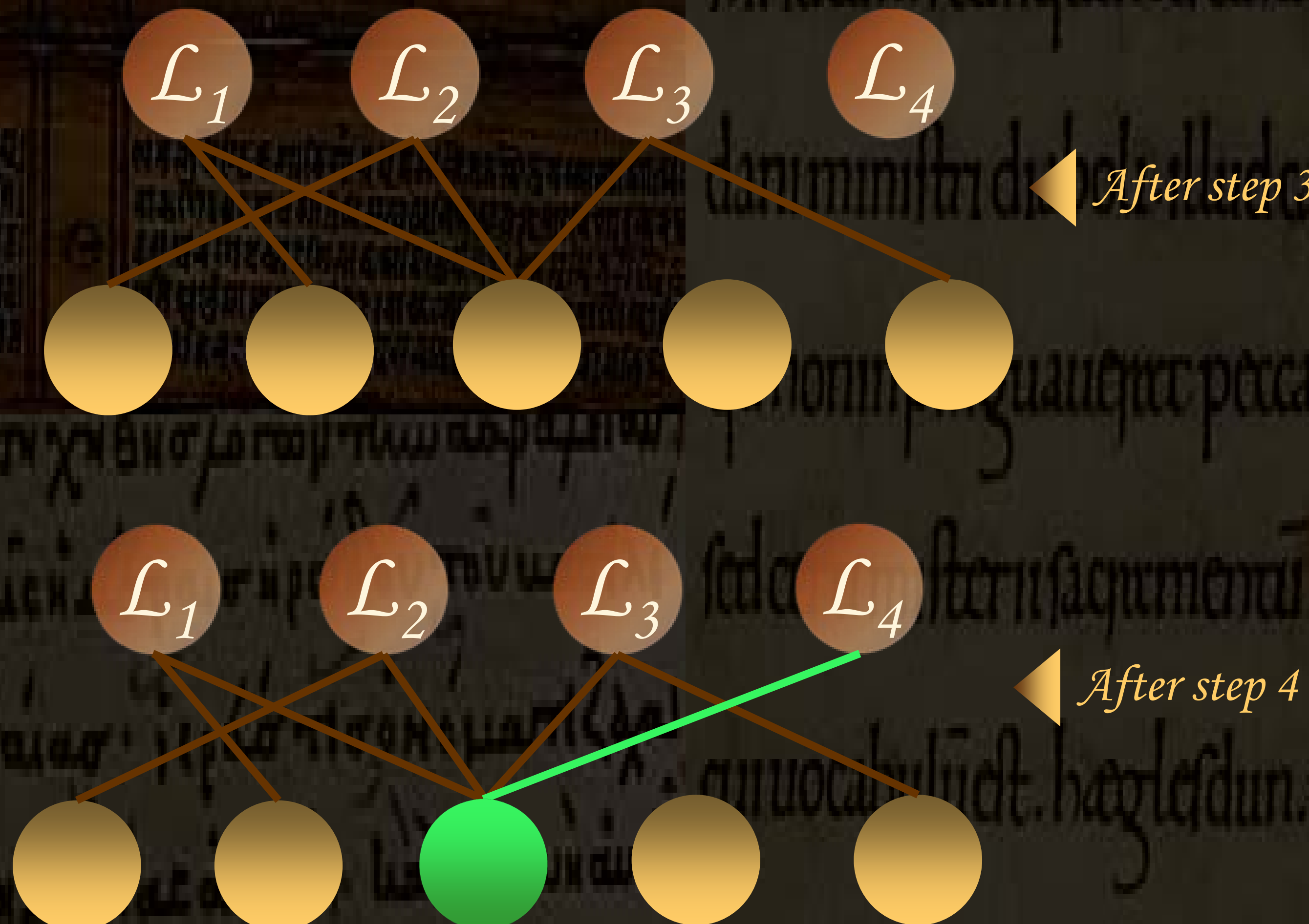


- Four models based on preferential attachment that incrementally attempt to explain the properties of PlaNet and PhoNet

Experimental Setup

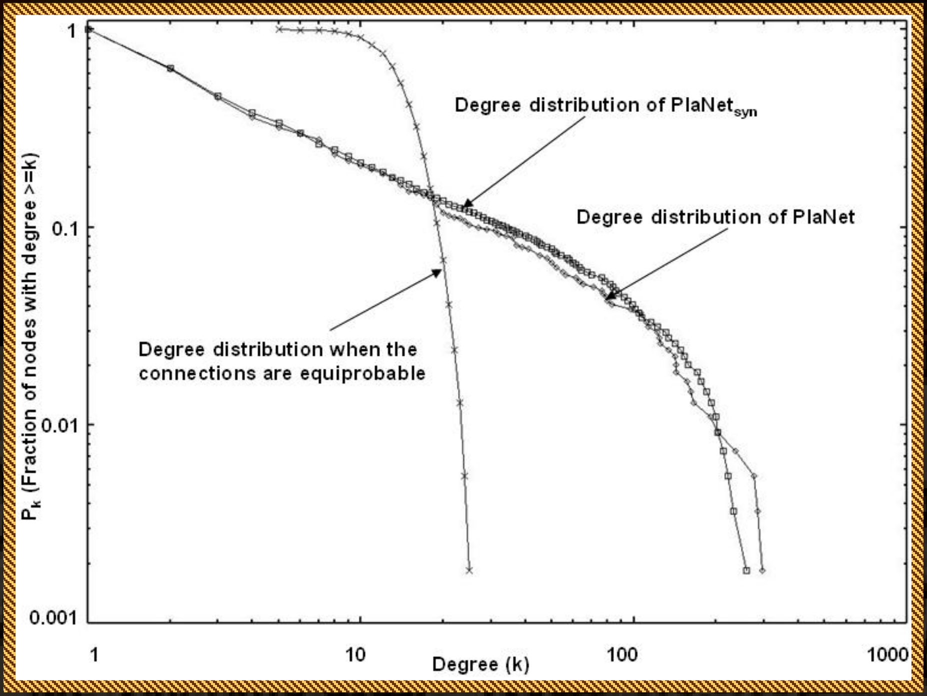
- Each model is a generative process (say P) conditioned by a set of parameters (say θ)
- P(θ) is a stochastic process governing the attachment of a language node to a consonant node, where the attachment is proportional to the current degree of the consonant node in the partially synthesized network
- Monte Carlo simulations of P(θ) are conducted to obtain many instances of synthesized networks and the network properties are averaged over all these instances
- The following four models attempt to incrementally match the different properties of the real networks (for certain choice of θ)

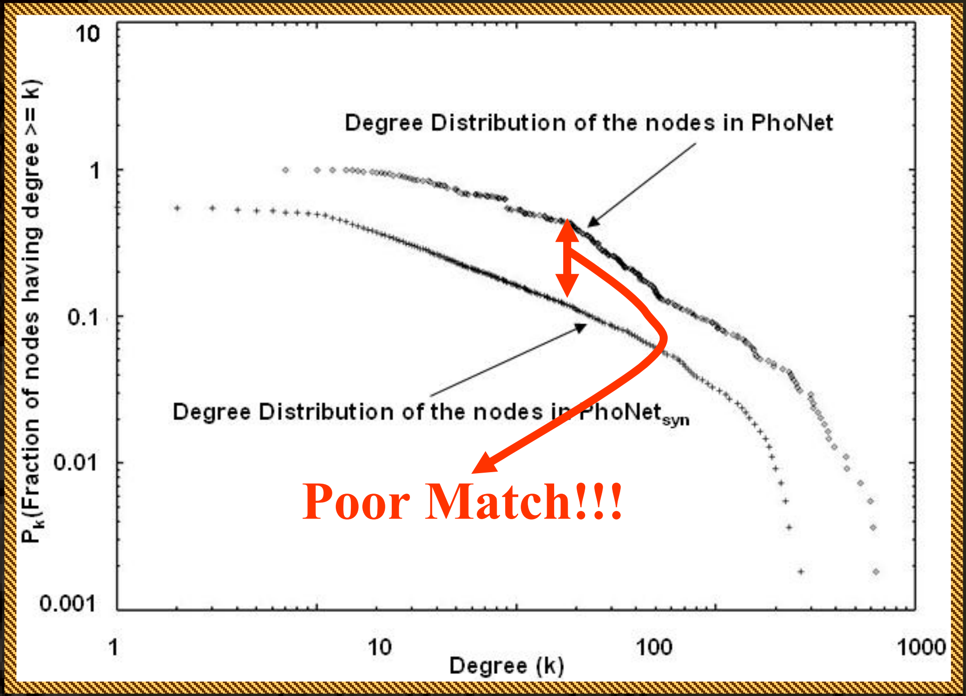
Model I: Preferential Attachment (PA)



Results

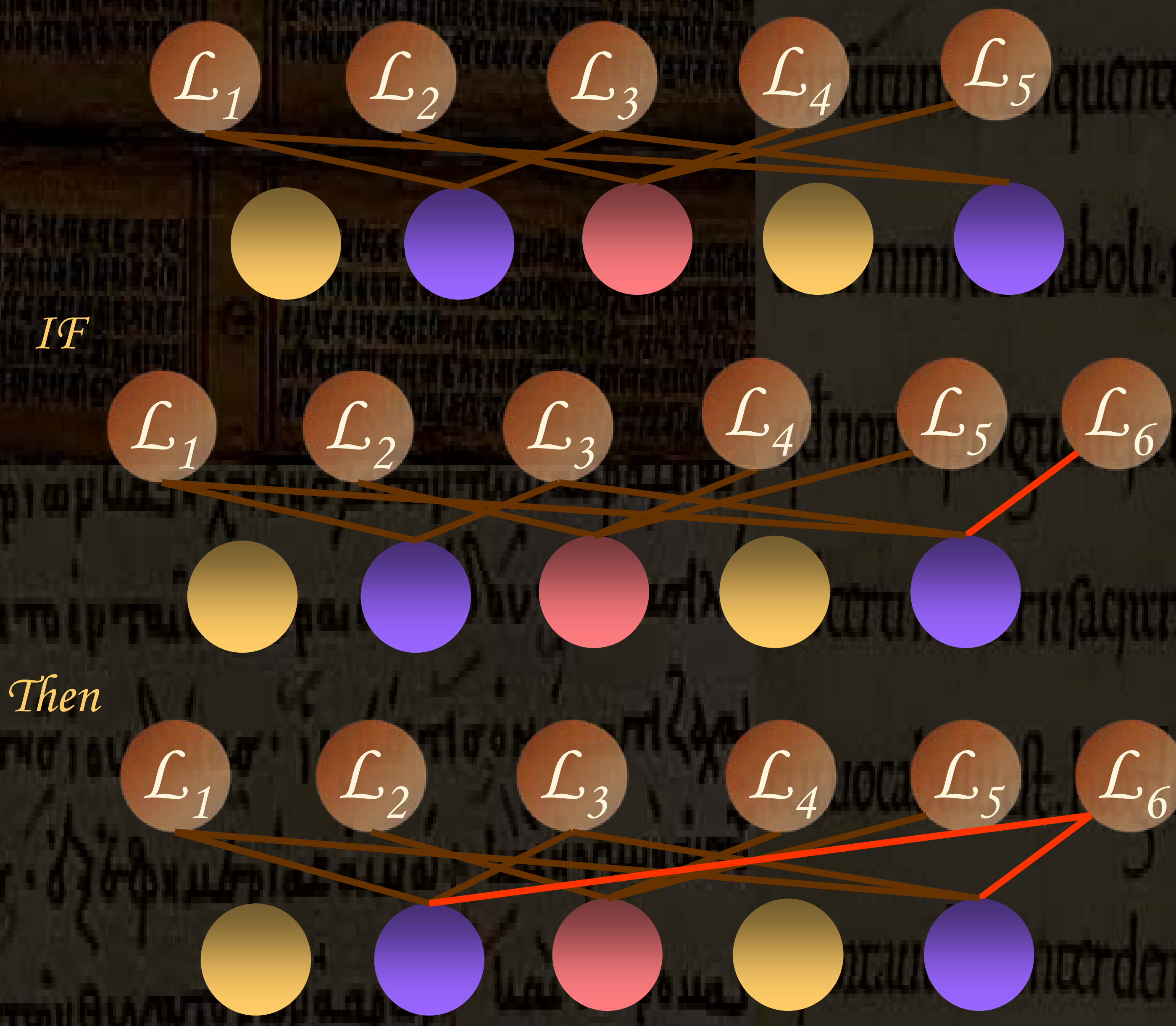
- Degree Distribution [ε=0.5,α=1.44]





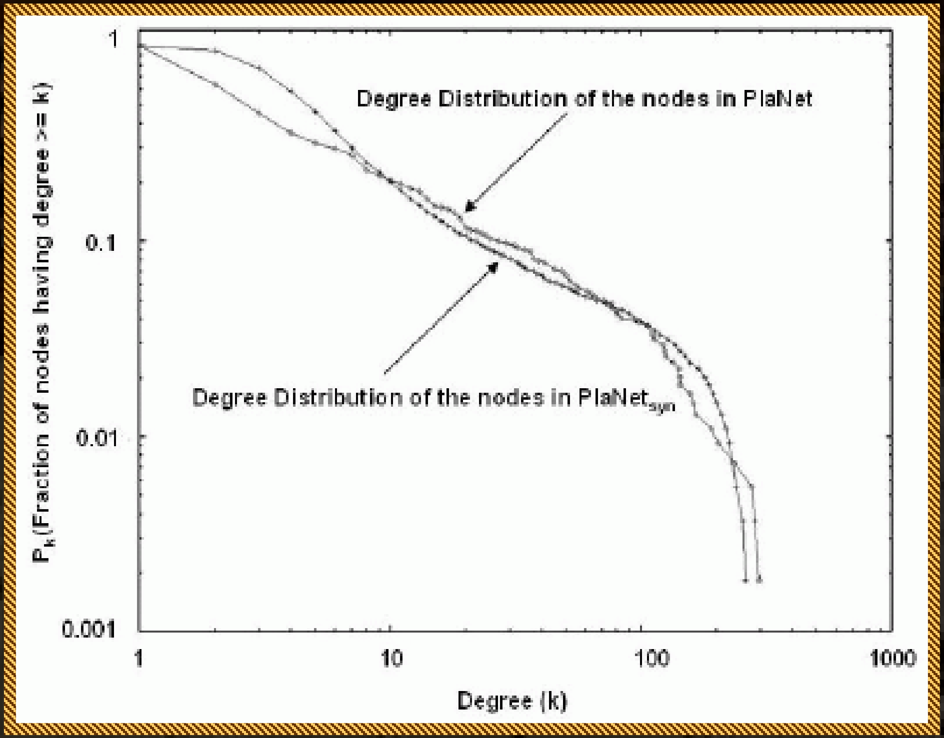
- Clustering Coefficient of PhoNet$_{syn}$ is 0.55 (for PhoNet it was 0.89) → Poor Match!!!

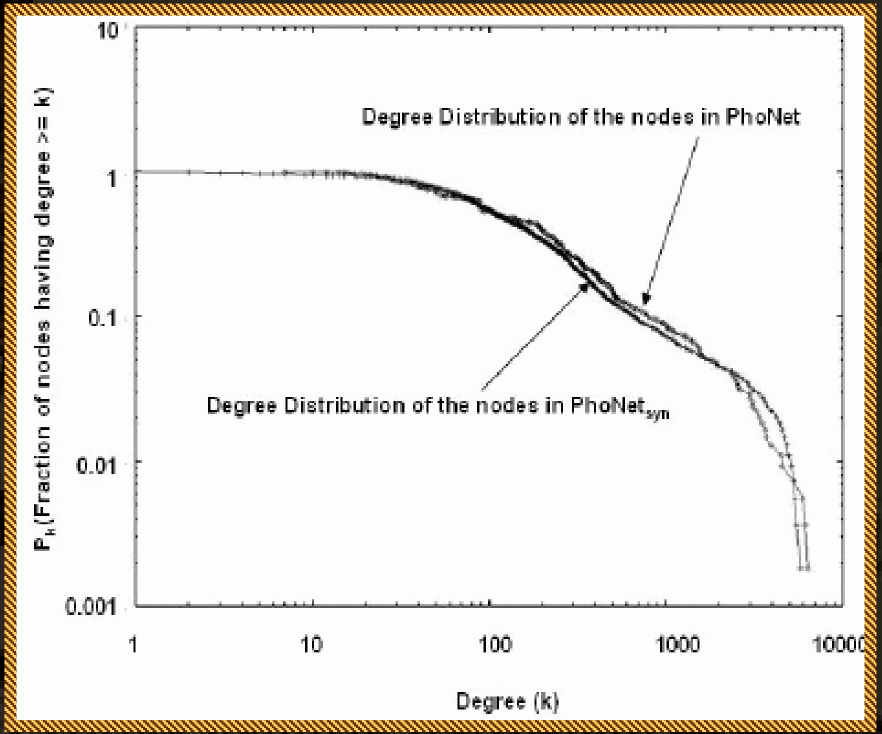
Model II: PA + Triad Formation



Results

- ➢Degree Distribution [ε=0.3,α=1.3,p_t=0.8]

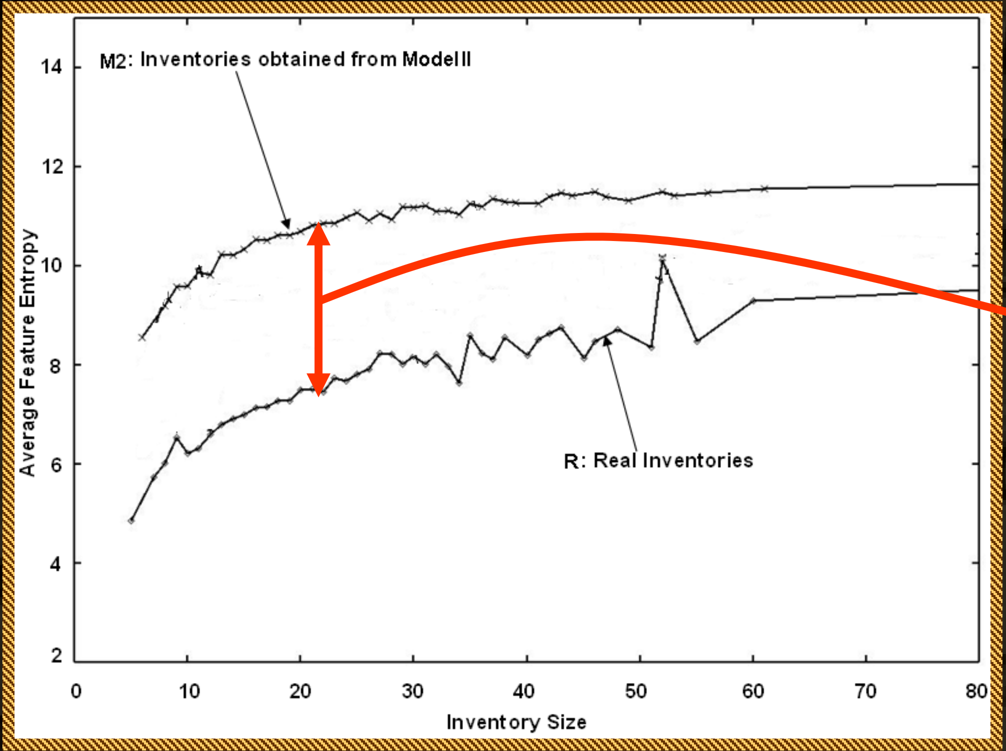




- ➢This time the clustering coefficient of PhoNet$_{syn}$ is 0.85 (within 4.5% of PhoNet)

What about Feature Entropy?

- Nodes are unlabeled in the synthesized networks
  - Sort the consonants of UPSID in the decreasing order of their frequency of occurrence and call this list of consonants *List*[1 · · · 541]
  - Sort the V_C nodes of PlaNet_{syn} in decreasing order of their degree and call this list of nodes *List'* [1 · · · 541],
  - $\forall_{1 \le i \le 541}$ *List'* [i] ← *List* [i]



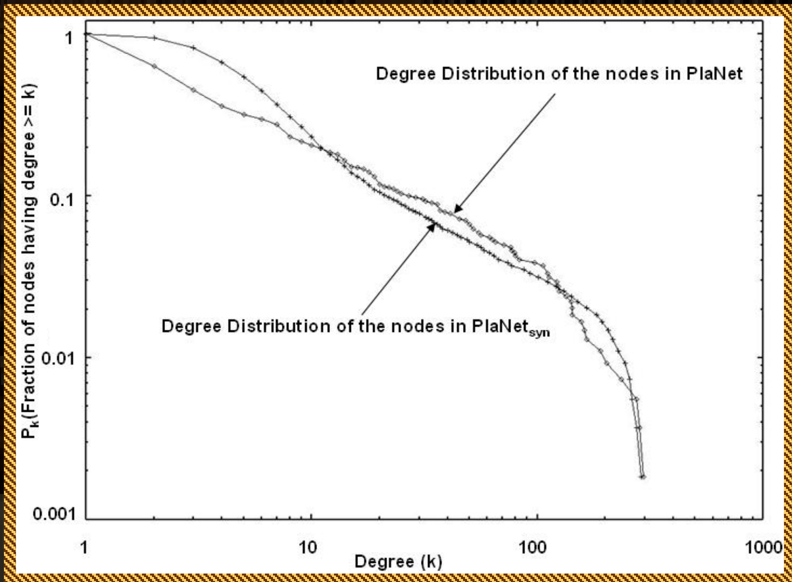
Huge Gap!! Can this be reduced?

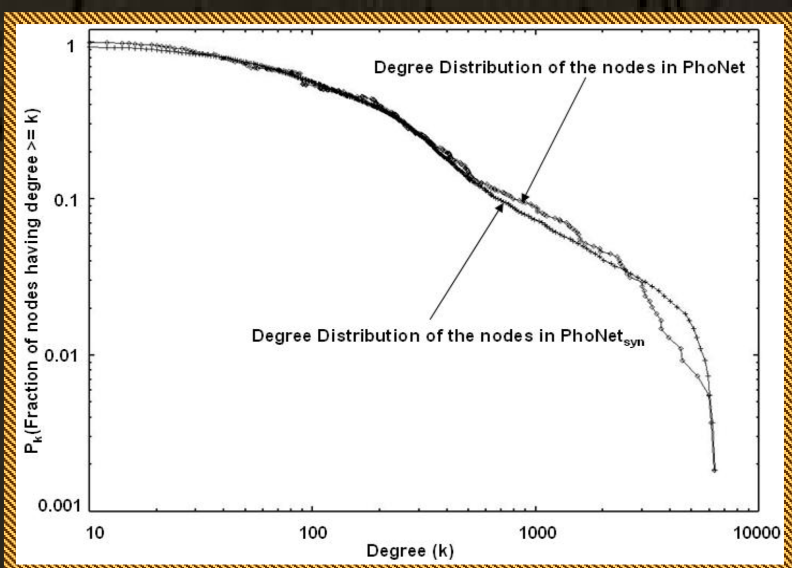
Model III: Feature-based Kernel

- Here we assume that nodes in V_C are labeled by their features, which are binary-valued. Hence, each node is a binary vector
- We define affinity between two nodes C_i and C_i' as
  - $A(C_i, C_i') = 1/D(C_i, C_i')$ where D is hamming distance
- The first consonant is chosen preferentially following *PA*
- Connection to any other consonant C_i' is based on
  - $(1\text{-}w)PA + wPr_{aff}(C_i, C_i')$ where
  - $Pr_{aff}(C_i, C_i') = A(C_i, C_i')/\sum_{\forall C_i'} A(C_i, C_i')$

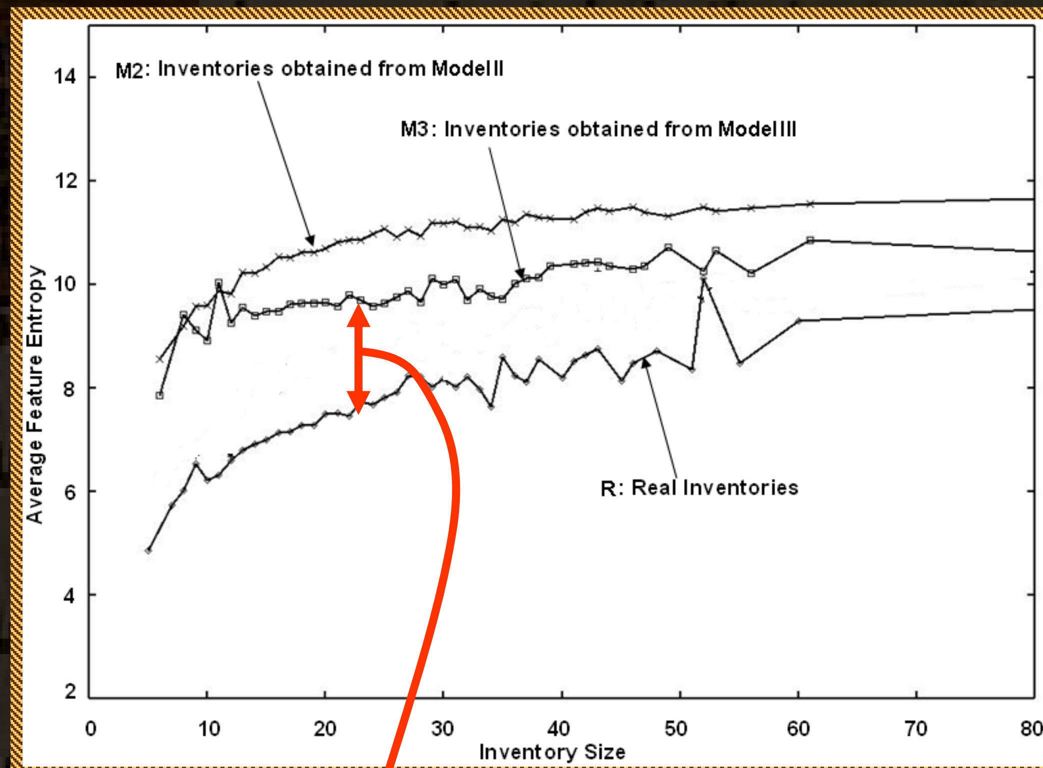
Results

[ε=0.3,α=1.6,w=0.2]





- Clustering coefficient is 0.84 (within 5.6% of PhoNet)



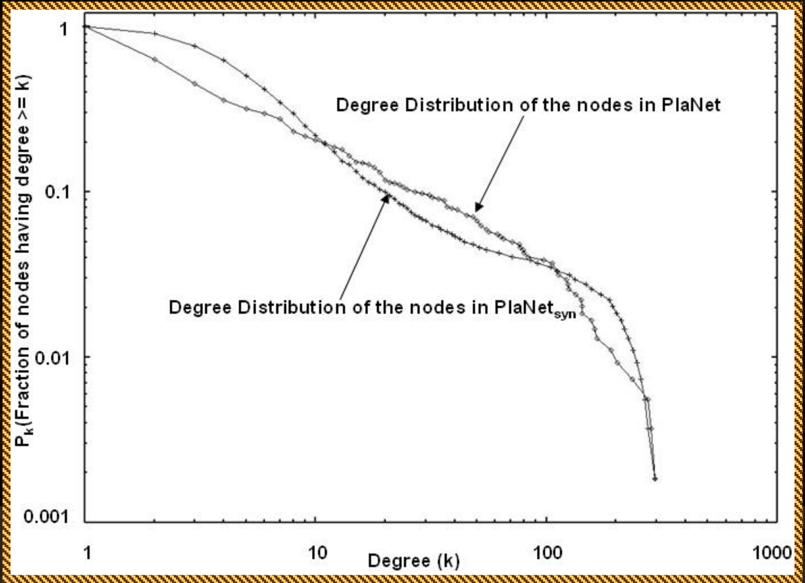
This time its better! But can we reduce the difference further?

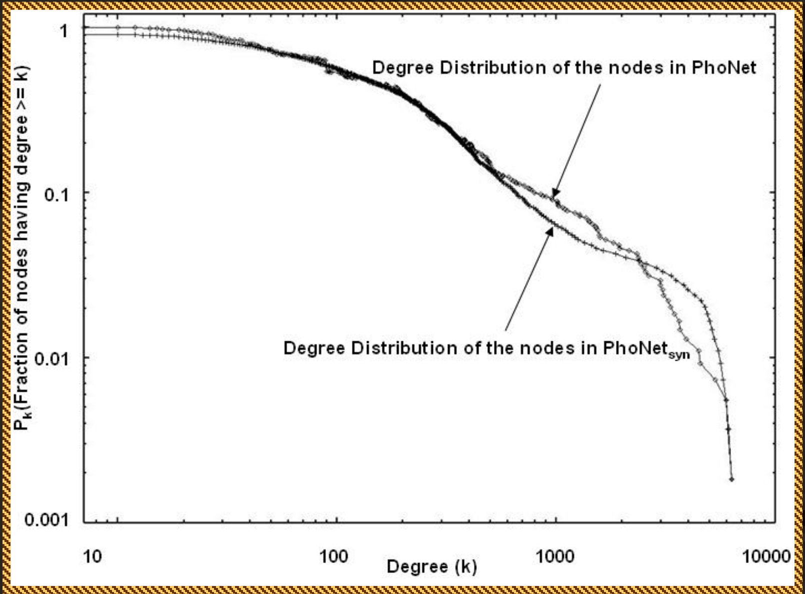
Model IV: Feature-based Kernel + Bootstrapping

- An initial bias towards the labeling prevalent in PlaNet
- Select 30 real inventories (around 10%) at random from UPSID and construct a PlaNet from them. Call this network the initial PlaNet$_{syn}$.
- The rest of the language nodes are incrementally added to this initial PlaNet$_{syn}$ using Model III.

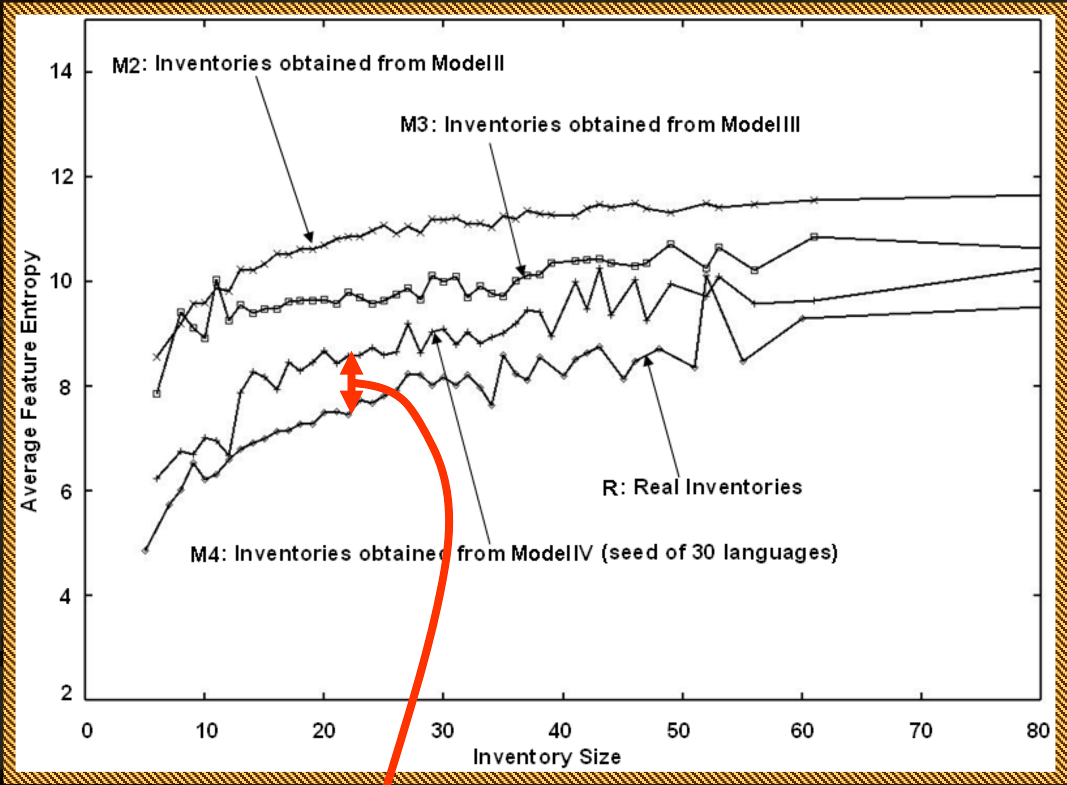
Results

[ε=0.3,α=1.35,*w*=0.15]





- ➢ Clustering coefficient is 0.83 (within 6.7% of PhoNet)



The more the number of languages used for bootstrapping, the better is the match.

Reflections

- Language is a constantly changing system and preferential attachment plays a significant role in this change.
  - 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).
  - The heterogeneity in the choice of the consonants manifests itself as preferential attachment.
  - Further, if a group of consonants largely co-occur in the languages of a generation then it is very likely that all of them get transmitted together in the subsequent generations
  - The groups of consonants are driven by feature economy

Reflections

- At the level of an individual, the significance of the models can be explained in terms of language acquisition
  - During *babbling* clear preferences of an infant are observed towards cross-linguistically frequent sounds (even though these are absent in the native language of the child) → This innate preference is captured by the preferential part of the models
  - 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 →This is modeled through the affinity of nodes based on features


To-DAH