# Language Variation: Convergence, Divergence, & Death

Robert Daland
Janet Pierrehumbert

Presented at the First Annual Workshop on Complex Systems & Language Workshop

University of Arizona
April 25-26, 2008

# Acknowledgements


- McDonnell foundation

- Celina Troutman
- Matt Goldrick
- Brady Clark
- Forrest Stonedahl
- Sara Solla

- NICO
- Phonatics

# Social coordination of grammars

- General property of population-level models
  - If agents try to imitate others, the population will converge
- Examples
  - Vowel systems (de Boer, 2000)
  - Word meanings (Cucker, Zhou, & Smale, 2004)
  - Word order (Niyogi, 2006)
  - *Bilingualism (Abrams & Strogatz, 2007)
- Problem: Empirically false

# Persistence of diversity

- Speakers are demonstrably in contact but differ in
  - Vowel system
    - British and American English
    - Southern and Midwestern American
  - Word meanings
    - British and American English
    - Southern and Midwestern American
  - Dominant language
    - Catalan/Spanish

- Theoretical challenge – keep socially-driven convergence, add persistence of diversity

# Plan

- Develop mathematical framework to analyze language community as dynamical system
  - Individual social and cognitive effects
- Quantify timescale(s) of convergence
  - Eigenvalue and Lyapunov number
  - Broad conditions for convergence
  - Interaction between social and cognitive timescale gives rise to linguistically important effects
- Social/structural simulations
  - Quantify effect of social/structural factors on language change

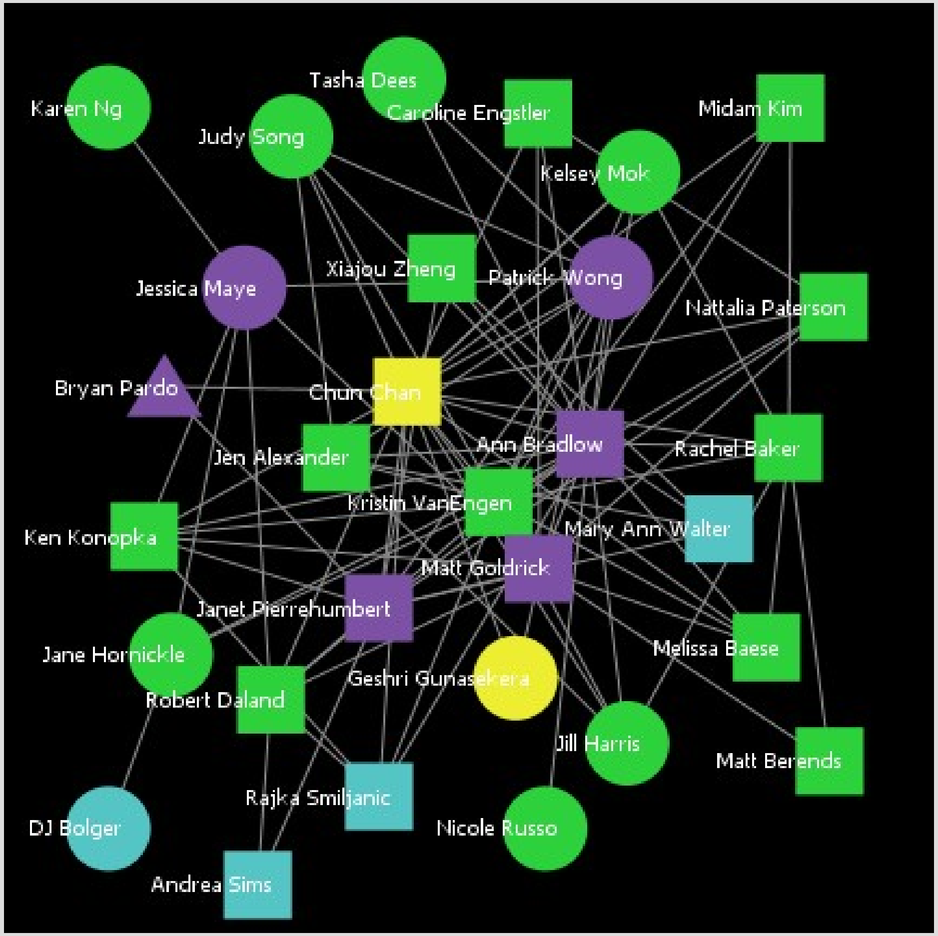
# Plan

- Develop mathematical framework to analyze language community as dynamical system
  - Individual social and cognitive effects
- Quantify timescale(s) of convergence
  - Eigenvalue and Lyapunov number
  - Broad conditions for convergence
  - Interaction between social and cognitive timescale gives rise to linguistically important effects
- Social/structural simulations
  - Quantify effect of social/structural factors on language change

# General assumptions

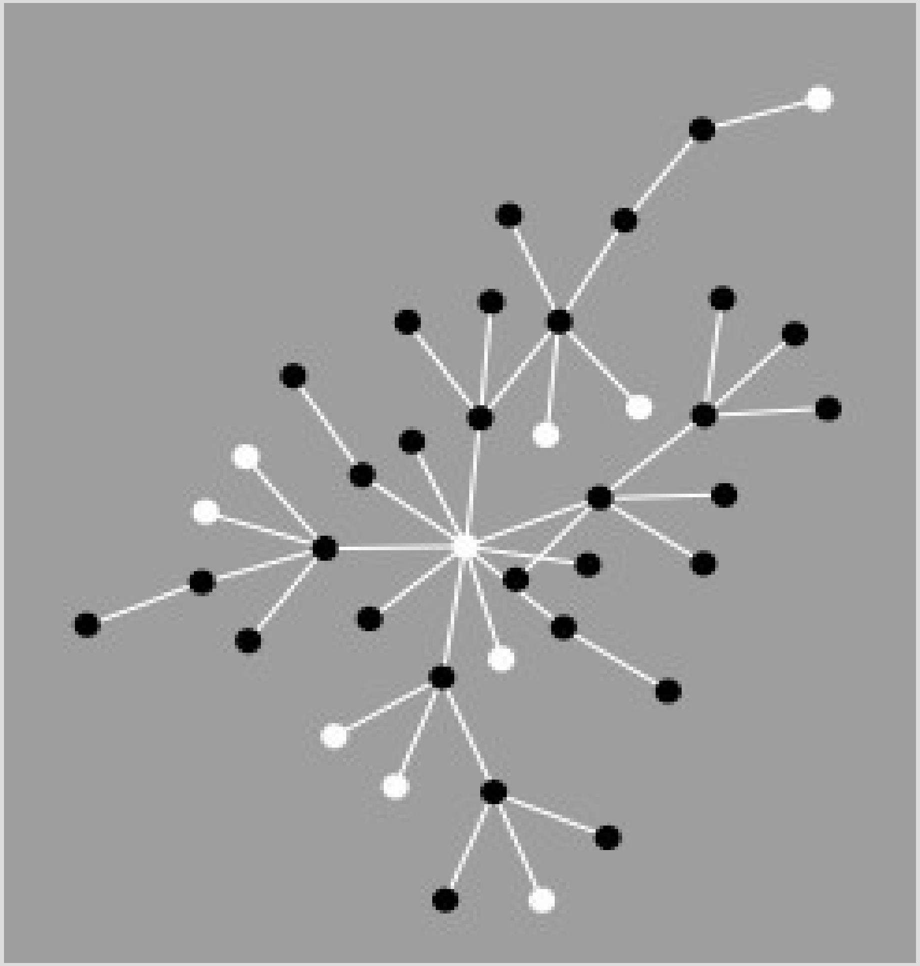
- Model community with "agents" connected by a social network
- Agents have a grammatical state
- Each time step connected agents converse
  - Talk: produce one output token according to current grammatical state
  - Listen: update grammatical state based on outputs of neighbors

# Social model




- Social networks
  - Many agents
  - Complicated relationships
  - Varying degrees of importance

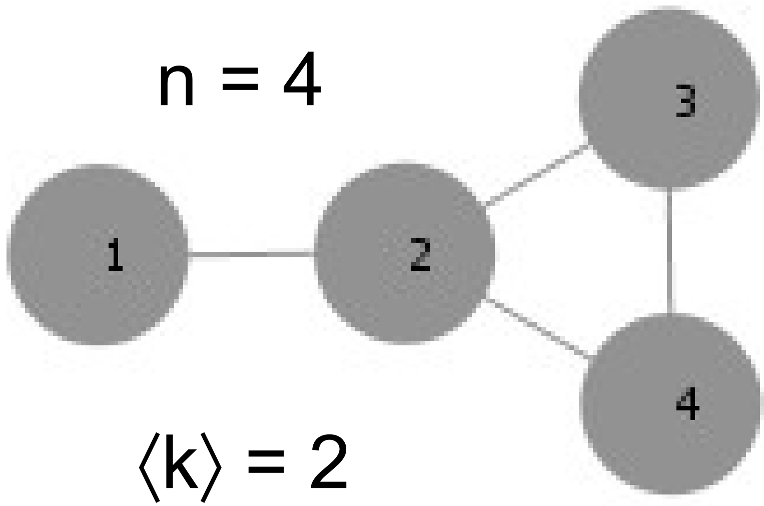
# Social model



- Social networks
  - Many agents
  - Complicated relationships
  - Varying degrees of importance
- Idealization
  - Few agents
  - One kind of relation
  - Uniform importance

# Social Model

$n = 4$




$\langle k \rangle = 2$

- Adjacency matrix
  - $A_{ij} = 1$ if *i* knows *j*
  - Symmetric

$$K = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

# Language model

- Production model
  - Categorical output: 0 or 1
  - Probability of each depends on grammar
  - Cognitive pressure: may bias production probabilities

$$p(out^t = 1) = f(g^t)$$

- Perception model
  - Productions are perceived correctly
  - Social pressure: Listener updates grammar to be more like neighbor's production

$$g^{t+1} = \alpha \cdot out_j^t + (1-\alpha) \cdot g^t$$

Learning/Memory rate

# Example

[movie 1 -- TGC]

# Analysis

- Simulations reveal the properties of a specific system
  - Easy to manipulate parameters and see effect
  - Not always clear why effects obtain

- Mathematical analysis gives insight into why the system behaves that way
  - Enables specific predictions about range of behaviors a system can (and cannot) exhibit

# Cast to dynamical system

- Let $\mathbf{g}^t = (g_1^t, g_2^t, \ldots, g_n^t)$ be vector of grammars
- Write update step for individual $g_i^t$
  - One conversation: $g_i^{t+1} = \alpha \cdot out_j^t + (1-\alpha) \cdot g_i^t$
  - All neighbors: $g_i^{t+1} = \alpha \cdot (\sum_j A_{ij} \cdot \langle out_j^t \rangle) + (1-k_i \cdot \alpha) \cdot g_i^t$
  - And since $\langle out_j^t \rangle = f(g_j^t)$...
- Collect into system equation:

$$\mathbf{g}^{t+1} = (\alpha A) \cdot f(\mathbf{g}^t) + (I - \alpha K) \cdot \mathbf{g}^t$$

# Plan

- Develop mathematical framework to analyze language community as dynamical system
  - Individual social and cognitive effects
- Quantify timescale(s) of convergence
  - Eigenvalue and Lyapunov number
  - Broad conditions for convergence
  - Interaction between social and cognitive timescale gives rise to linguistically important effects
- Social/structural simulations
  - Quantify effect of social/structural factors on language change

# “No cognitive bias” case

- Learners are faithful to input distribution
  - e.g. speech errors (Goldrick & Larson, in press)
- Formalized here by making f(g) = g
  - Left-hand side *g* represents the perceived input distribution
  - Right-hand side *g* represents the output distribution
  - “Faithful” means output distribution matches input distribution
- No cognitive bias
  - In the sense that physical/cognitive constraints do not bias the speaker toward one alternative

# Pseudo-Markov chain

$$\mathbf{g}^{t+1} = (\alpha A)\cdot f(\mathbf{g}^{t}) + (I - \alpha K)\cdot\mathbf{g}^{t}$$

becomes

$$\mathbf{g}^{t+1} = (\alpha A)\cdot\mathbf{g}^{t} + (I - \alpha K)\cdot\mathbf{g}^{t}$$

$$\mathbf{g}^{t+1} = (I + \alpha(A - K))\cdot\mathbf{g}^{t}$$

$$\mathbf{g}^{t+1} = M\cdot\mathbf{g}^{t}$$

- M is row-stochastic (all row-sums are 1)
  - $\text{row}_i$-sum of A = #nbrs of *i* = $\text{row}_i$-sum of K

# Properties

- Application of Perron-Frobenius Theorem
  - M has a unique dominant eigenvalue $\lambda_1 = 1$ and a corresponding eigenvector $\nu_1 = (1, 1, ..., 1)$
  - $\lim_{t\to\infty} \mathbf{g}^t = \lim_{t\to\infty} M^t \cdot \mathbf{g}^0 = c\nu_1$ $\quad c = (\sum_i g_i^0)/n$
- All agents converge to same stochastic grammar *c* (stable variation)
  - Consistent with social-network-theoretic literature

# Rate Analysis

- Convergence guaranteed, practical time horizon, not
  - How long does it take to reach equilibrium state?

- Measure *half-life*
  - amount of time required to half present “distance” from equilibrium

$$\tau_{social} = -1/\lg |\lambda_2|$$

# Cognitive bias case

- Learners are somehow unfaithful to input statistics
  - Characterized abstractly: locus of effects in $f$
- Bistability: Assume that 0-type and 1-type are stable
  - If whole population is 0/1, it will stay there
- Monotonicity: Assume $f$ is monotonic
  - Higher rate of 1's in the input never causes speaker to produce lower rate of 1's

# Mean-field approximation

- Theoretical construct: average agent
  - Has average grammatical state, average number of neighbors, etc...

$$\mathbf{g}^{t+1} = (\alpha A)\cdot f(\mathbf{g}^{t}) + (I - \alpha K)\cdot \mathbf{g}^{t}$$

becomes

$$\langle \mathbf{g}^{t+1} \rangle = \alpha \langle k \rangle \cdot f(\langle \mathbf{g}^{t} \rangle) + (1 - \alpha \langle k \rangle)\cdot \langle \mathbf{g}^{t} \rangle$$

$$\langle \mathbf{g}^{t+1} \rangle = F(\langle \mathbf{g}^{t} \rangle)$$

# Properties

- F inherits properties from *f*

- Bistability and monotonicity imply that
  - There are attracting fixed points at 0 and 1
  - There is a repelling fixed point $0 < r < 1$
  - There are no other fixed points

- Solution:
  - If $\langle \mathbf{g}^0 \rangle < r$: $\lim_{t \to \infty} \langle \mathbf{g}^t \rangle = 0$
  - If $\langle \mathbf{g}^0 \rangle > r$: $\lim_{t \to \infty} \langle \mathbf{g}^t \rangle = 1$

# Rate Analysis

- Nonlinear analogue to 2nd eigenvalue: Lyapunov exponent
  - $\lg \lambda_T(F, \langle g \rangle^0) = \langle \lg F'(\langle g \rangle^t) \rangle$ where $T = \lceil \tau_{coupling} \rceil$
- Time constant defined analogously

$$\tau_{cognitive} = -1 / \lg \lambda_T(F, \langle g \rangle^0)$$

In principle F can be estimated with psycholinguistic studies, e.g. Hudson, Kam, & Newport (2005).

# Conditions for convergence

- Convergence under no cognitive bias
  - Faithful to input statistics
- Convergence under mean-field assumption
  - Valid when “average” agent is representative
  - Certainly occurs when every agent has same grammar
  - E.g. when social pressure drives everyone together faster than cognitive pressure could drive them apart

Guaranteed when $\tau_{social} < \tau_{cognitive}$

# Example of divergence

movie 2 – TGC splitting

# Dialect splitting

- Can be observed when $\tau_{cognitive} > \tau_{social}$
- Interaction between cognitive structure of speaker/listeners and social structure of language community

# Plan

- Develop mathematical framework to analyze language community as dynamical system
  - Individual social and cognitive effects
- Quantify timescale(s) of convergence
  - Eigenvalue and Lyapunov number
  - Broad conditions for convergence
  - Interaction between social and cognitive timescale gives rise to linguistically important effects
- Social/structural simulations
  - Use rate computation to investigate social/structural factors in language change

# General idea

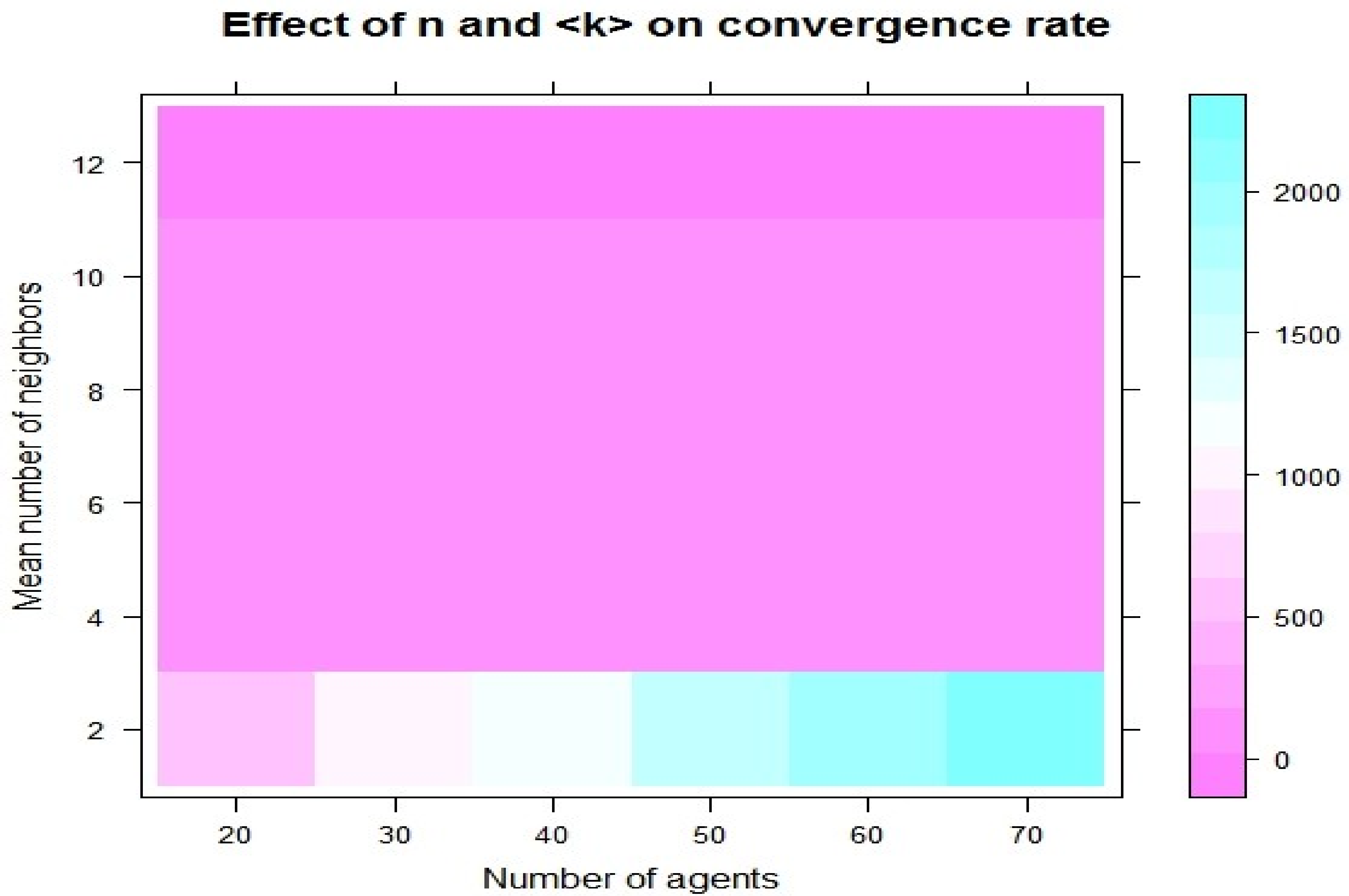
- Rate computation
  - Abstract computation
  - hard to interpret in terms of specific linguistic phenomena
  - But **easy** to interpret relationship between two computations

- Use rate computation to compare (models of) different language communities

# Question I

- How sensitive is the model to particular choices of n and k?
  - Languages are robust across orders-of-magnitude variation in n (50 to 50 million)
  - Should be a regime where model exhibits this same behavior
- Simulations
  - Vary n and $\langle k \rangle$
  - For each (n, $\langle k \rangle$) pair create 100 preferential attachment networks
  - Calculate average $\tau_{social}$

# Simulations: Convergence Rate



# Discussion

- Phase transition
  - In the very low-k regime, convergence is sensitive to population size
  - Elsewhere, it is basically insensitive to total population size

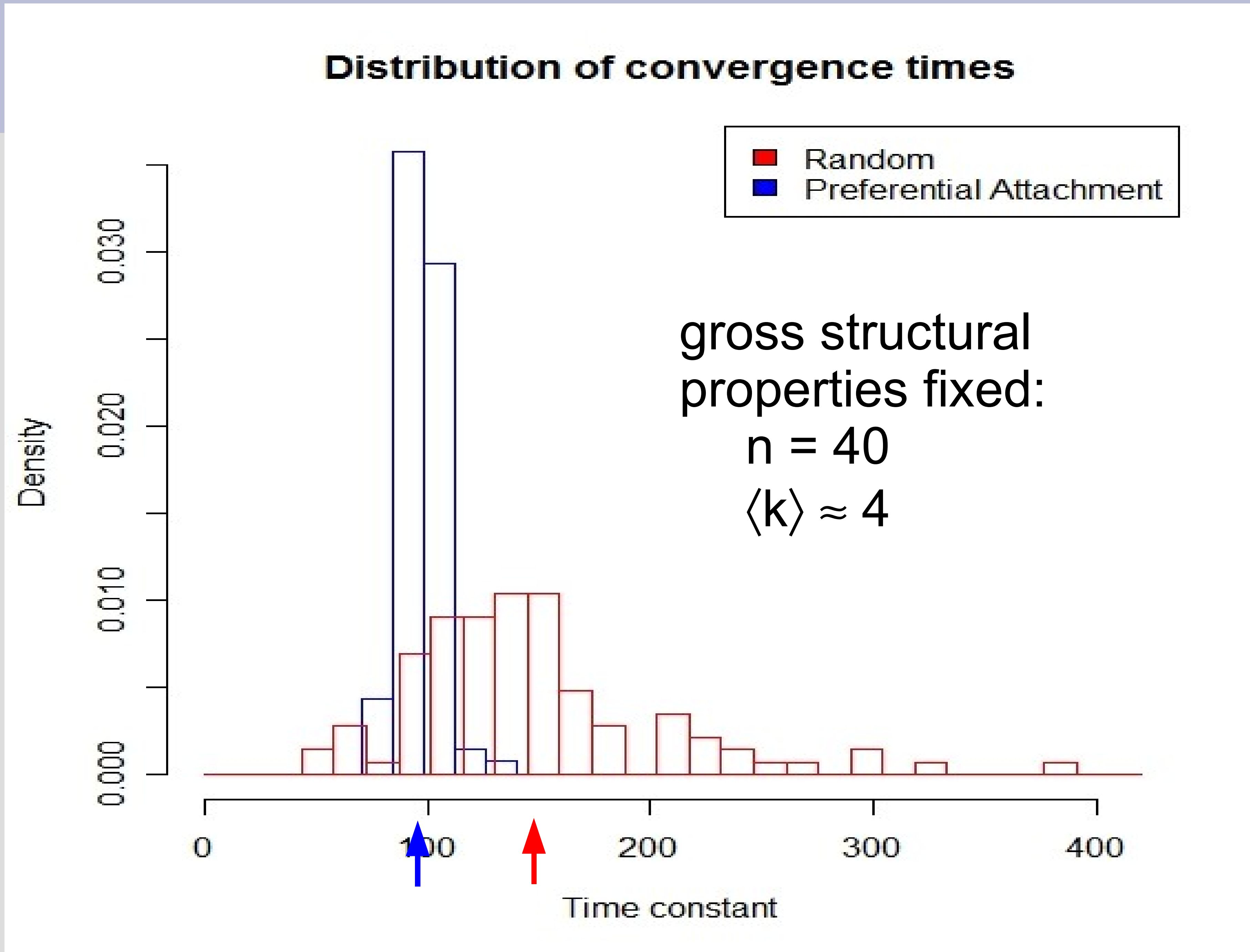
- Low k-regime
  - Population has a tree structure, in which most agents only have a *single* neighbor

- The interesting action is in k
  - k is a gross structural property
  - Do finer-grained properties matter?

# Question II: Open/closed society

- Sociolinguists have long speculated that social structure impacts language change
- Investigate here by varying social structure
- Network structural types
  - Random – egalitarian link distribution
  - Preferential-attachment – few with many, many with few

# Simulation




gross structural properties fixed:

$n = 40$

$\langle k \rangle \approx 4$

# Discussion

- Different social structures exhibit different language dynamics
  - Generally faster for preferential-attachment network
  - Much lower variability on preferential-attachment network

- Offers quantitative support for sociolinguistic intuition

# Summary

- Dynamical systems theory
  - Tool to conceptualize and analyze language as a socially-embedded process
- Interplay of social and cognitive factors gives rise to linguistically important phenomena
  - Nonlinearity can amplify inter-community variation while reducing intra-community variation
- Rate analysis gives insight on social factors in language change
  - Hierarchical societies transmit grammars efficiently
  - Population density more important than size

# Open questions and issues I

Heterogeneity

- In resistance to input from others
- In the production function
- In nature and function of links

# Heterogeneity in link weighting

- Social model assumes that interaction always makes agents converge

- Not always true in the wild
  - E.g. some speakers may try to maintain a distinct social identity by creating/maintaining a distinct linguistic identity

- Allow negative weighting in social interactions
  - I.e. relax assumption that adjacency matrix is non-negative
  - N.b. causes most of the theory to no longer apply

[movie 3 -- stars 'n sneetches]

# Open questions and issues II

- Assumption of perfect perception
  - Pearl & Weinberg (2007) account for English shift from OV to VO word order by assuming that learners discard ambiguous data

# Open questions and Issues III

- Model validation
  - Tying abstract variables of model to observables
  - e.g. Hudson, Kam, & Newport (2005)
  - Qualitative/Comparative predictions

# Dialect-splitting

- Stable variation of some linguistic variable X
  - $\tau_{cognitive} >> \tau_{social} > 0$
- Dialect splitting on some linguistic variable Y
  - $0 < \tau_{cognitive} < \tau_{social}$
- Two dialects of a common origin which have split along Y should have the same stable variation X
  - optional-*that* should be equally probable in British and American English

# Thank you

- Questions/Comments?