Modeling Population Structure and Language Change in the St. Louis Corridor

Jordan Kodner
University of Pennsylvania

NWAV 46, November 6, 2017
University of Wisconsin

Outline

- History of the Northern Cities Shift in the St. Louis Corridor
- Frameworks for Population-Level Change
- Modeling Change in the Corridor

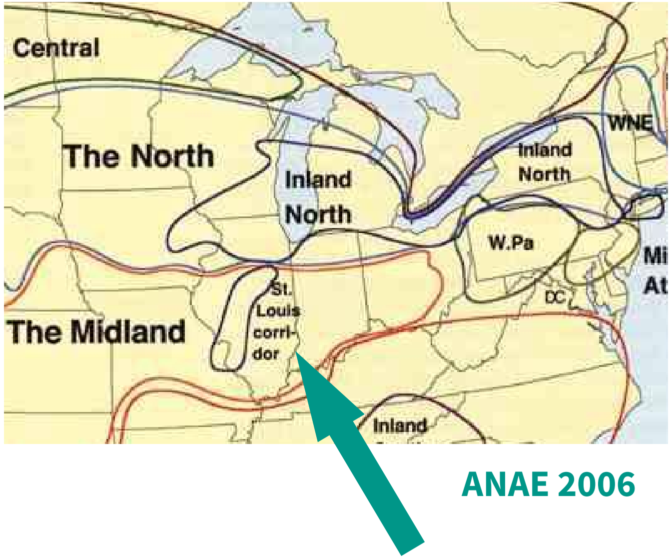
Slides Available Here:

ling.upenn.edu/~jkodner

History of the NCS in the St. Louis Corridor

The St. Louis Corridor

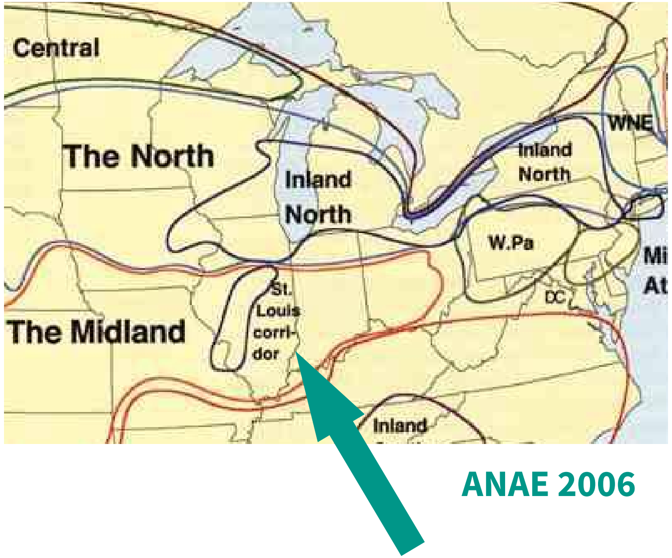
- Dialect region within Midlands between Chicago and St. Louis
- And Inland North “island”



ANAE 2006

The St. Louis Corridor

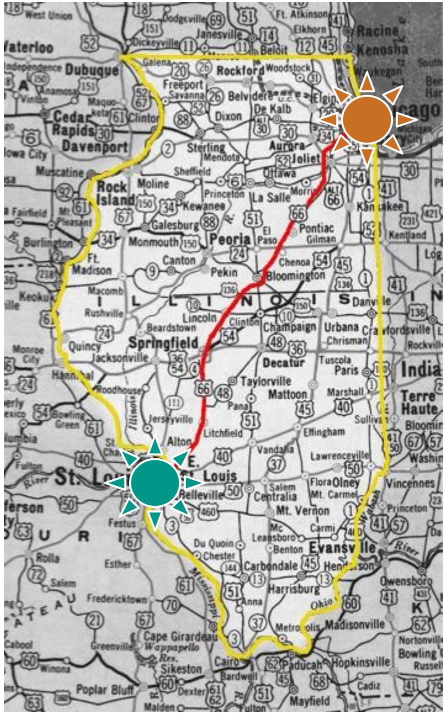
- Dialect region within Midlands between Chicago and St. Louis
- And Inland North "island"
- The Northern Cities Shift has advanced and retreated there



ANAE 2006

Shape of the Corridor

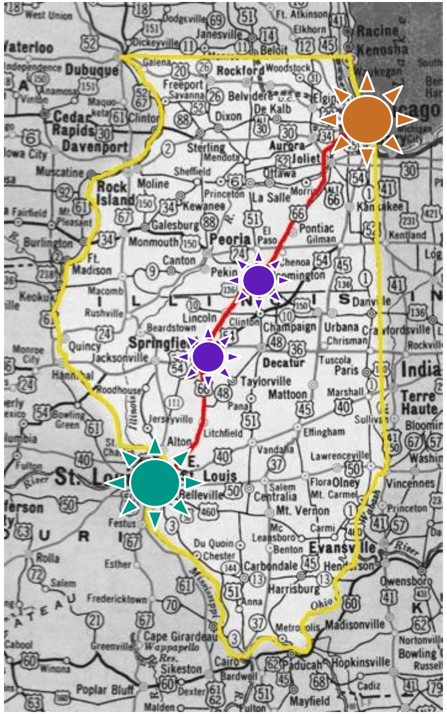
- **Follows Old Route 66 from outside Chicago to St. Louis**



Friedman 2014

Shape of the Corridor

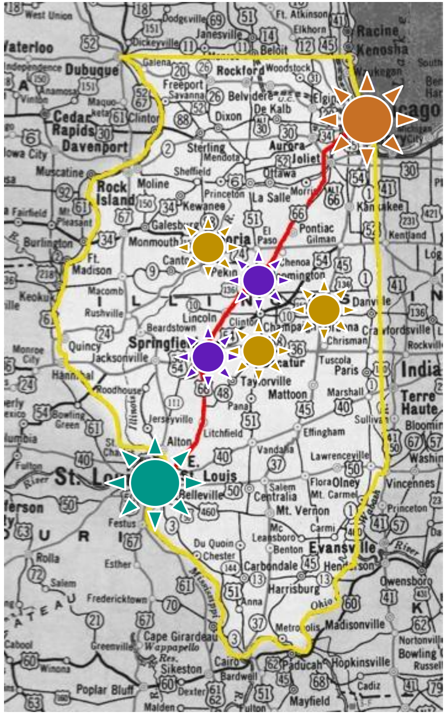
- Follows Old Route 66 from outside Chicago to St. Louis
- Route 66 passed through Springfield and Bloomington,



Friedman 2014

Shape of the Corridor

- Follows Old Route 66 from outside Chicago to St. Louis
- Route 66 passed through Springfield and Bloomington,
- And near Decatur, Peoria, Champaign,
- And many small towns



Friedman 2014

Route 66

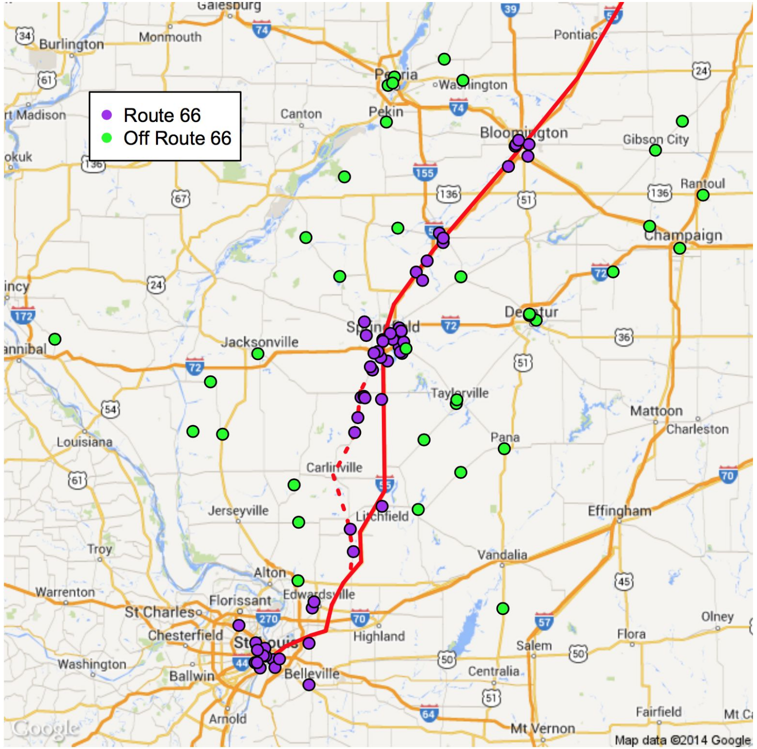
- **Construction began in 1926**
- **Replaced a series of unpaved roads and canals**
- **Connected the main streets of towns along its path**

Route 66

- Construction began in 1926
- Replaced a series of unpaved roads and canals
- Connected the main streets of towns along its path
- Superseded by I-55 in 1957
- Decommissioned in 1985

NCS in the Corridor

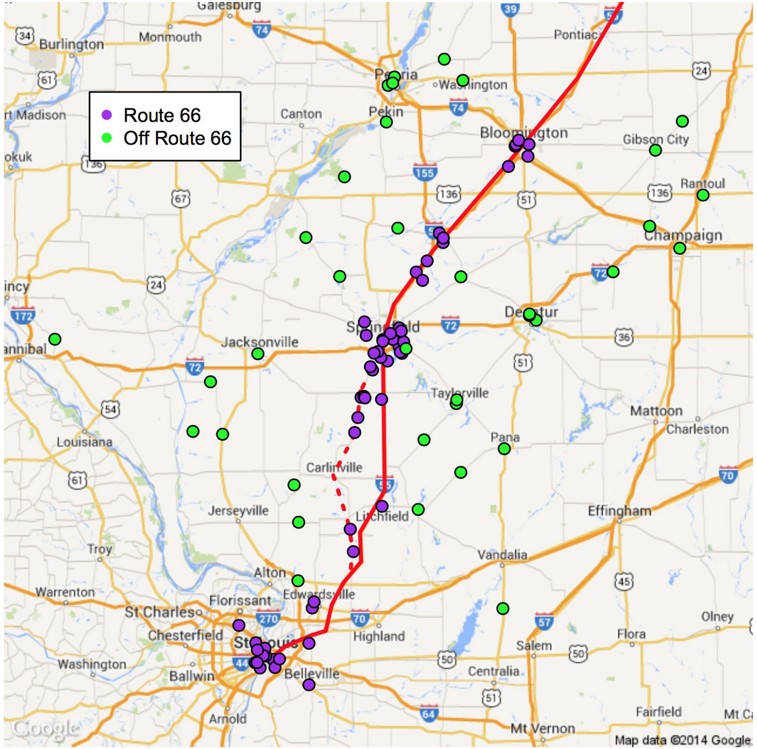
- **NCS entered the Corridor via Route 66 during the Great Depression**
  (Only period with net migration out of Chicago into smaller cities)



Friedman 2014

NCS in the Corridor

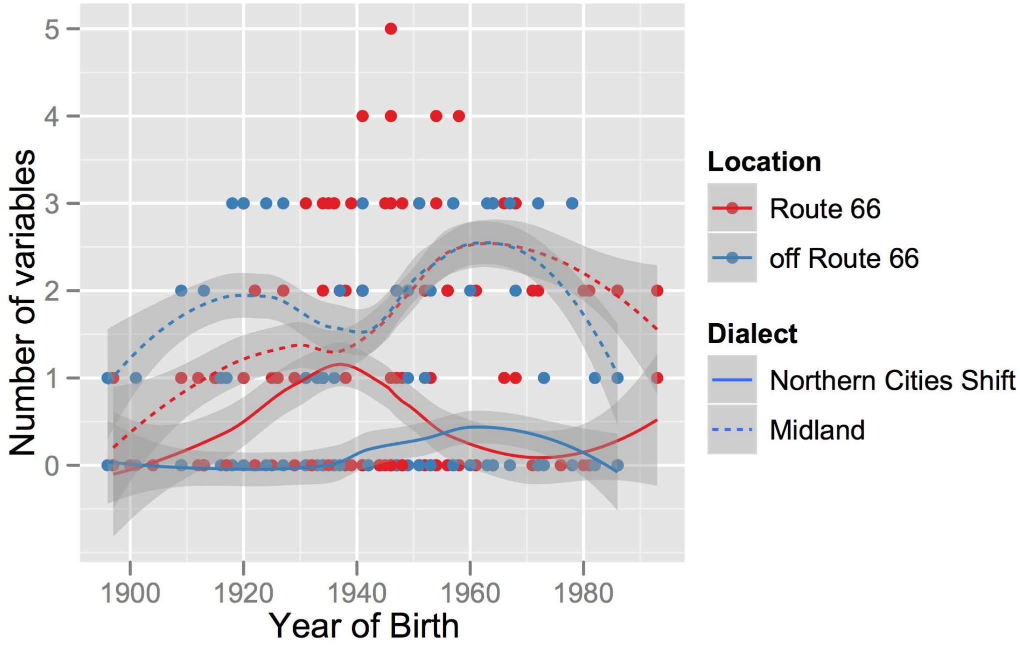
- NCS entered the Corridor via Route 66 during the Great Depression
  (Only period with net migration out of Chicago into smaller cities)
- NCS observed first in "on-route" cities, then in "off-route" cities
- Has since largely receded



Friedman 2014

The St. Louis Corridor

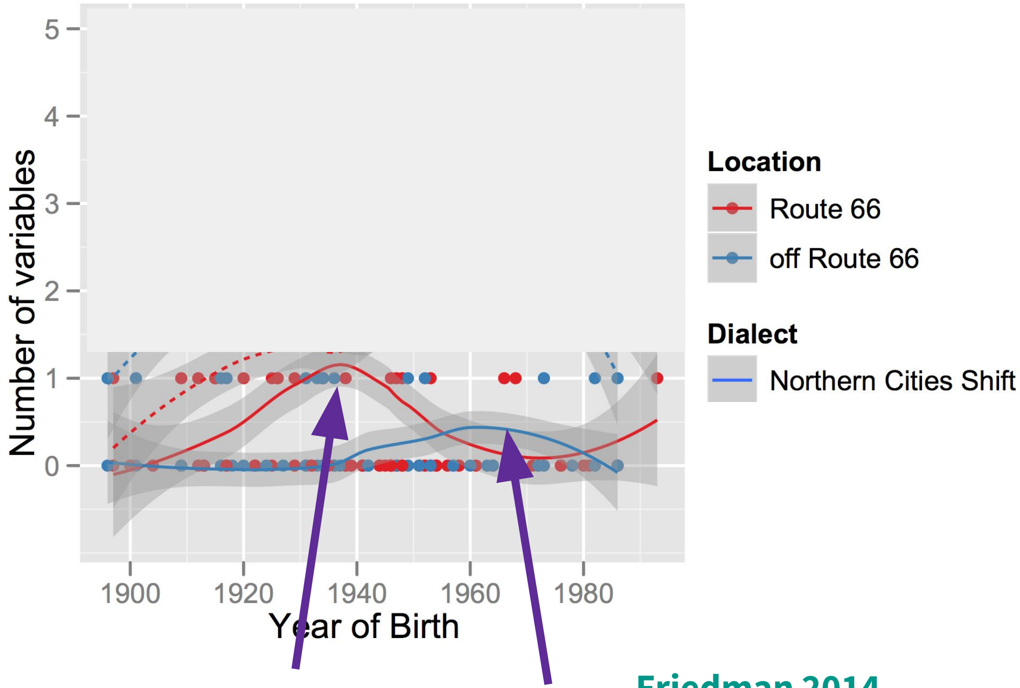
- **Path of change is different On-Route and Off-Route**
  - **NCS peaks first On-Route**
  - **NCS peaks higher On-Route**
  - **Peaks Off-Route about one generation later**



Friedman 2014

The St. Louis Corridor

- **Path of change is different On-Route and Off-Route**
  - NCS peaks first On-Route
  - NCS peaks higher On-Route
  - Peaks Off-Route about one generation later



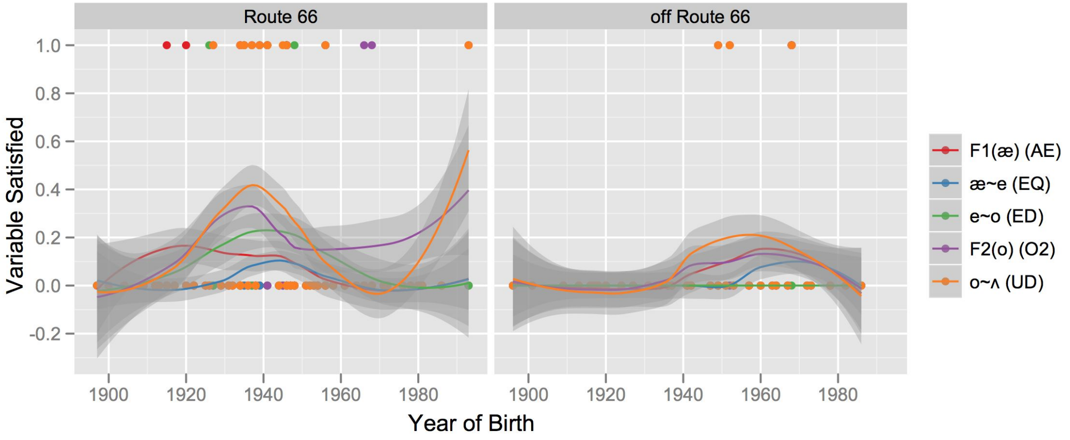
Friedman 2014

The St. Louis Corridor

- **Path of change is different On-Route and Off-Route**
  - **NCS peaks first On-Route**
  - **NCS peaks higher On-Route**
  - **Peaks Off-Route about one generation later**
- **Similar path for all variables**

On-Route

Off-Route



Friedman 2014

The Hypothesis (Friedman 2014)

- **Adult speakers imported the NCS to the Corridor in the 1930s**
- **It was transmitted to learners in On-Route communities**
- **These diffused it to nearby towns Off-Route**
- **Where is it was acquired by learners about a generation later**

The Hypothesis (Friedman 2014)

- Adult speakers imported the NCS to the Corridor in the 1930s
- It was transmitted to learners in On-Route communities
- These diffused it to nearby towns Off-Route
- Where is it was acquired by learners about a generation later

- Historical data is highly suggestive of this course of events
- Is there a way to test it without a time machine?

The Hypothesis (Friedman 2014)

- Adult speakers imported the NCS to the Corridor in the 1930s
- It was transmitted to learners in On-Route communities
- These diffused it to nearby towns Off-Route
- Where is it was acquired by learners about a generation later

- Historical data is highly suggestive of this course of events
- Is there a way to test it without a time machine?
- Yes! We can simulate it

Modeling Population-Level Change

Why Simulate Change?

- **We have lots of data on historical change and change in progress - evidence**

Why Simulate Change?

- **We have lots of data on historical change and change in progress - evidence**
- **We have logically derived theories of change - evidence**

Why Simulate Change?

- **We have lots of data on historical change and change in progress - evidence**
- **We have logically derived theories of change - evidence**
- **We have small-scale focused laboratory studies - evidence**

Why Simulate Change?

- We have lots of data on historical change and change in progress - evidence
- We have logically derived theories of change - evidence
- We have small-scale focused laboratory studies - evidence
- But we cannot test population-level change in the lab - missing evidence

It would be nice to test cause and effect directly.

Why Simulate Change?

- **We have lots of data on historical change and change in progress - evidence**
- **We have logically derived theories of change - evidence**
- **We have small-scale focused laboratory studies - evidence**
- **But we cannot test population-level change in the lab - missing evidence**

It would be nice to test cause and effect directly.

Simulation provides that outlet.

A useful tool in computational biology, epidemiology, … geology, etc.

Three Classes of Framework

1. **Concrete Frameworks**
2. **Network Frameworks**
3. **Algebraic Frameworks**

Three Classes of Framework

1. Concrete Frameworks

- **Individual agents on a grid moving randomly and interacting (e.g., Harrison et al. 2002, Satterfield 2001, Schulze et al. 2008, Stanford & Kenny 2013)**

\+ Gradient interaction probability for free

\+ Diffusion is straightforward

\- Not a lot of control over the network

\- Thousands of degrees of freedom -> should run many many times -> slow

\- Unclear how to include a learning model

Three Classes of Framework

1. **Concrete Frameworks**
2. **Network Frameworks**
   - **Speakers are nodes in a graph, edges are possibility of interaction (e.g., Baxter et al. 2006, Baxter et al. 2009, Blythe & Croft 2012, Fagyal et al. 2010, Minett & Wang 2008, Kauhanen 2016)**

   \+ **Much more control over network structure**

   \+ **Easy to model concepts from the sociolinguistic lit. (e.g., strong/weak)**

   \- **Nodes only interact with immediate neighbors -> slow and less realistic?**

   \- **Practically implemented as random interactions between neighbors -> same problem as #1**

Three Classes of Framework

1. **Concrete Frameworks**
2. **Network Frameworks**
3. **Algebraic Frameworks**

- **Expected outcome of all interactions is calculated directly (e.g., Abrams & Stroganz 2003, Baxter et al. 2006, Minett & Wang 2008, Niyogi & Berwick 1997)**

\+ **Less reliance on random processes -> faster and more direct**

\+ **Clear how to insert learning models into the framework**

\- **No network structure! Always implemented over perfectly mixed populations**

What We Use

- **An algebraic model operating on network graphs**

What We Use

- An algebraic model operating on network graphs
  - No random process in the core algorithm
  - Fast and efficient
  - Models language change in social structures

The best of both worlds!

Vocabulary for this Talk

Different research traditions, Different vocabularies

L: That which is transmitted

Language ≈ Variable ≈ *Lect ≈ E-Language

G: That which generates/describes/distinguishes L

That which is learned/influenced by L

Grammar ≈ Variant ≈ I-Language

The Model

Language change is a two step loop

1. **Propagation: calculate how L spread**
2. **Acquisition: calculate how G are learned**

The Model

Language change is a two step loop

1. **Propagation: calculate how L spread**
2. **Acquisition: calculate how G are learned**

If this were a linear chain,

$$L_0 \rightarrow G_1 \rightarrow L_1 \rightarrow G_2 \rightarrow L_2 \rightarrow \ldots \rightarrow L_n \rightarrow G_{n+1} \rightarrow \ldots$$

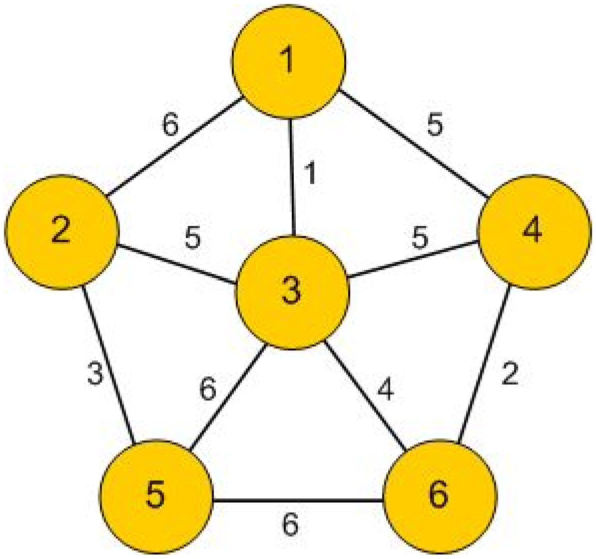
Mathematical Description

[REDACTED]

Propagation

Network Structure

- **Nodes**
  - **How many people are there?**
  - **How are people clustered? Socially or geographically?**
  - **Do people migrate?**

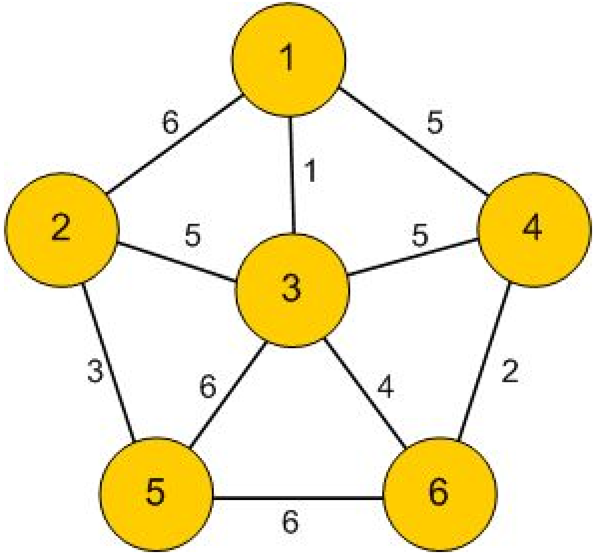


forum.qt.io

Propagation

Network Structure

- **Nodes**
  - How many people are there?
  - How are people clustered? Socially or geographically?
  - Do people migrate?
- **Edges**
  - Are interactions bidirectional?
  - Are interactions equal? By likelihood, frequency, or social valuation?
  - Can the mode of interaction change over time?

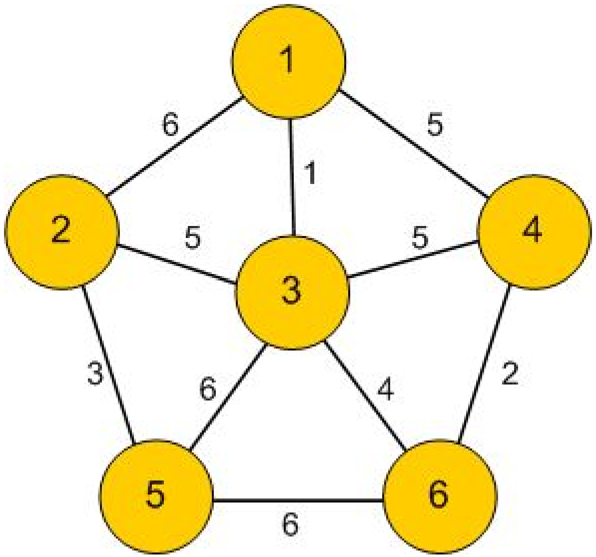


forum.qt.io

Propagation

Network Structure

- **Nodes**
  - How many people are there?
  - How are people clustered? Socially or geographically?
  - Do people migrate?
- **Edges**
  - Are interactions bidirectional?
  - Are interactions equal? By likelihood, frequency, or social valuation?
  - Can the mode of interaction change over time?
- **Replacement**
  - Are we modeling large scale (generations) or small scale (older/younger siblings) change?
  - Does the network grow or shrink?



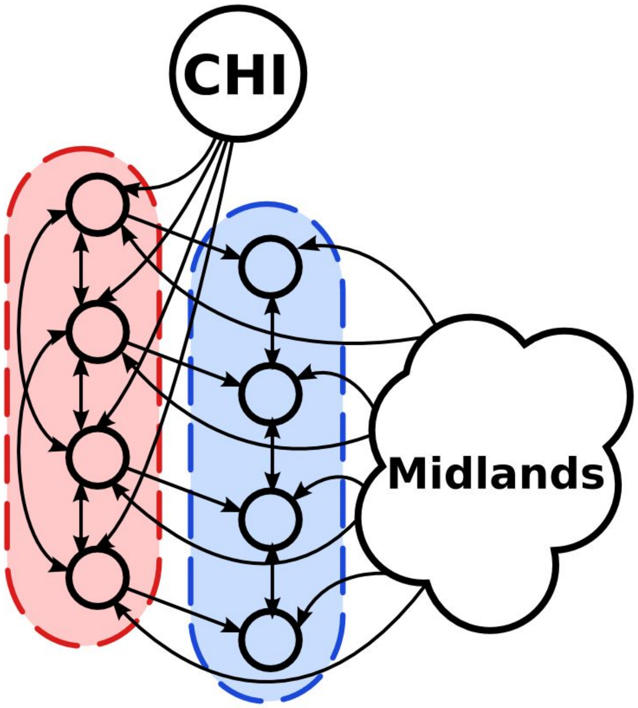
forum.qt.io

Modeling Change in the Corridor

Modelling the Corridor: Network Structure

Community Types:

- Midlands (1; “background”)
- Chicago (1)
- On-Route (19)
- Off-Route (19)



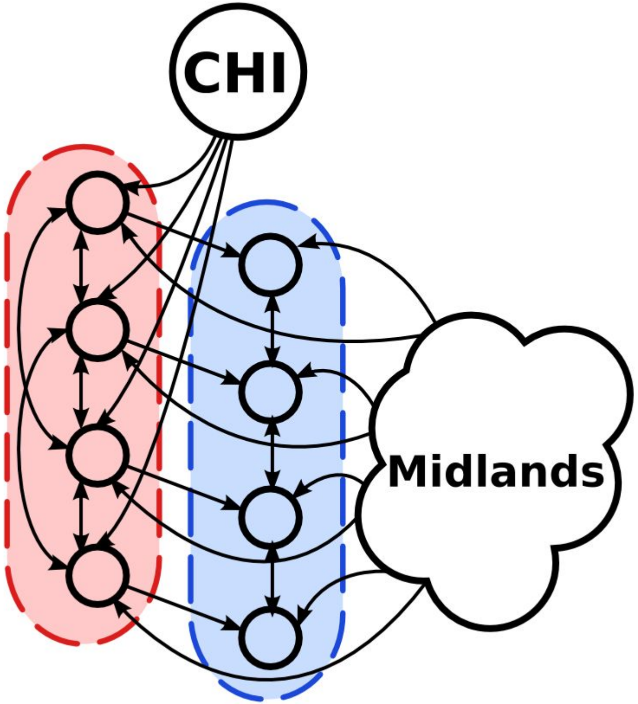
Modelling the Corridor: Network Structure

Community Types:

- Midlands (1; "background")
- Chicago (1)
- On-Route (19)
- Off-Route (19)

Connections:

- Midlands to all On-Route and Off-Route
- Chicago to all On-Route
- On-Route to two adjacent On-Route
- On-Route to one adjacent Off-Route
- Off-Route to one adjacent Off-Route



Modelling the Corridor: History

- **Vary a single parameter: Direction of movement to On-Route communities**
- **In order to test the Great Depression hypothesis**

Modelling the Corridor: History

- **Vary a single parameter: Direction of movement to On-Route communities**
- **In order to test the Great Depression hypothesis**
- **It would be too “easy” if we could vary multiple parameters**
    - Movement Off-Route
    - Strength of connections between On-Route and Off-Route
    - Strength of connections between On/Off-Route and Chicago/Midlands
    - Advantage of NCS
    - Etc.
- **And the results would be less meaningful**

Modelling the Corridor: History

- **Vary a single parameter: Direction of movement to On-Route communities**
- **In order to test the Great Depression hypothesis**
- **It would be too "easy" if we could vary multiple parameters**
  - Movement Off-Route
  - Strength of connections between On-Route and Off-Route
  - Strength of connections between On/Off-Route and Chicago/Midlands
  - Advantage of NCS
  - Etc.
- **And the results would be less meaningful**

Still important! Just not the focus of the current study...

Modelling the Corridor: History

- **Vary a single parameter: Direction of movement to On-Route communities**
- **In order to test the Great Depression hypothesis**

Stage 1 - 5 iterations

No migration (speaker interaction only)

Stage 2 - 20 iterations

2% migration from Chicago to On-Route "Great Depression"

Stage 3 - 75 iterations

2% migration from Midlands to On-Route "Post-Depression"

Modelling the Corridor: The Variable

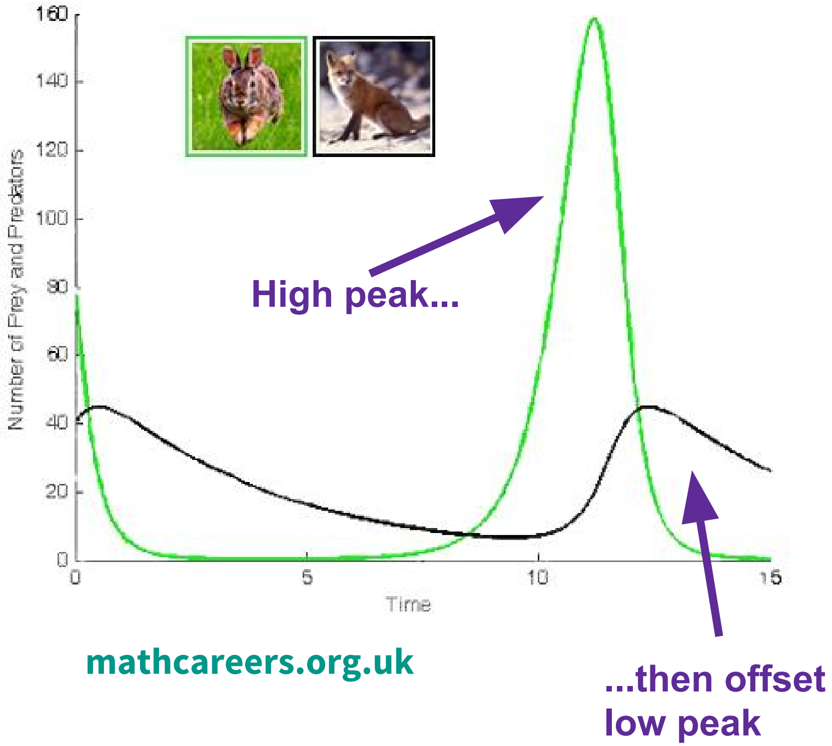
- **Treating the NCS as a single binary variable subject to competing grammars**
- **Community Variable Distributions:**
    - **Chicago fixed at 100% NCS+**
    - **Midlands fixed at 100% NCS-**
    - **On/Off-Route begins 100% NCS- but is allowed to vary**

Modelling the Corridor: The Variable

- **Treating the NCS as a single binary variable subject to competing grammars**
- **Community Variable Distributions:**
  - **Chicago fixed at 100% NCS+**
  - **Midlands fixed at 100% NCS-**
  - **On/Off-Route begins 100% NCS- but is allowed to vary**
- **Tested as neutral, slightly advantaged, and heavily advantaged change**

Two-Compartment Systems

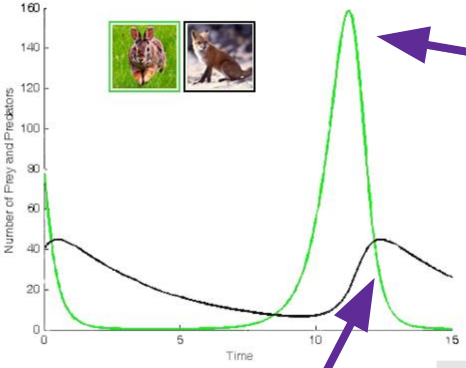
- **A type of dynamical system arising in in electrical engineering, medicine, chemistry, ecology...**



mathcareers.org.uk

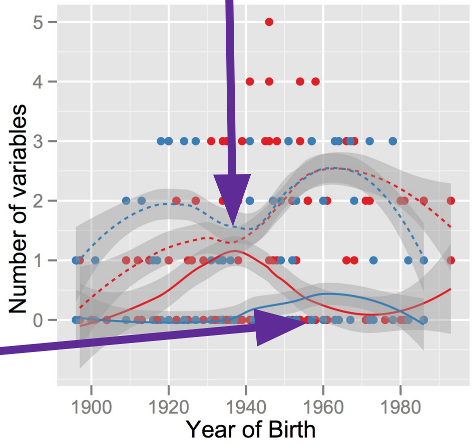
Two-Compartment Systems

- A type of dynamical system arising in in electrical engineering, medicine, chemistry, ecology...and linguistics!
- Here, On-Route and Off-Route are the compartments
- And the time for variable propagation is the delay



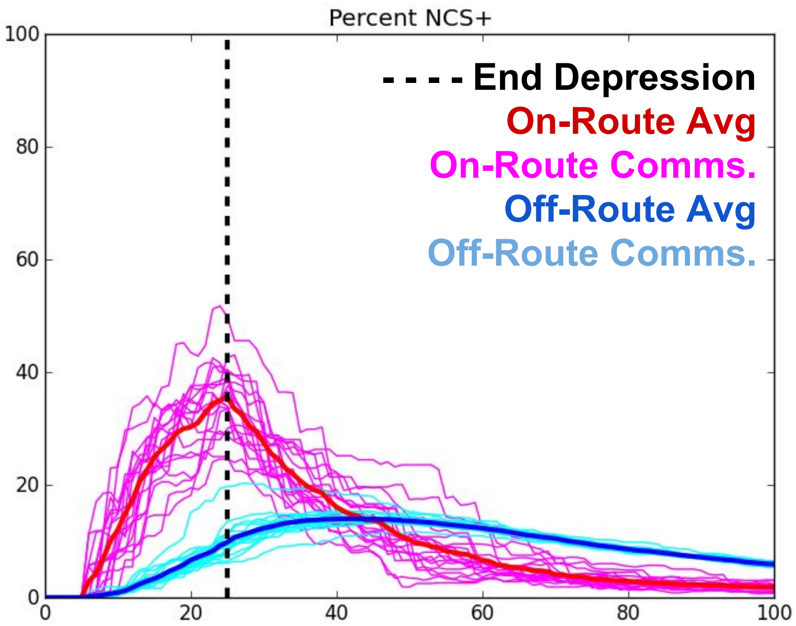
High peak...

...then offset low peak



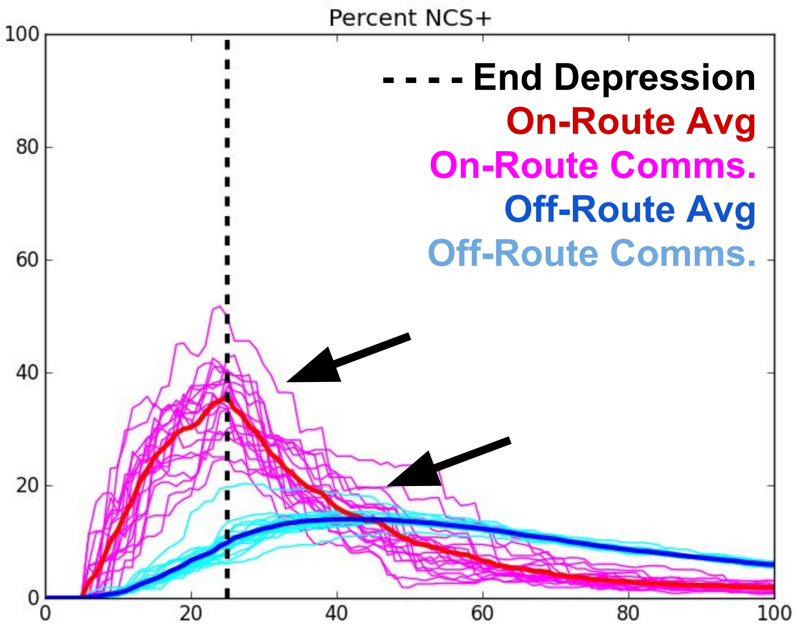
Results: Neutral Change

- **A classic two-compartment pattern arises**



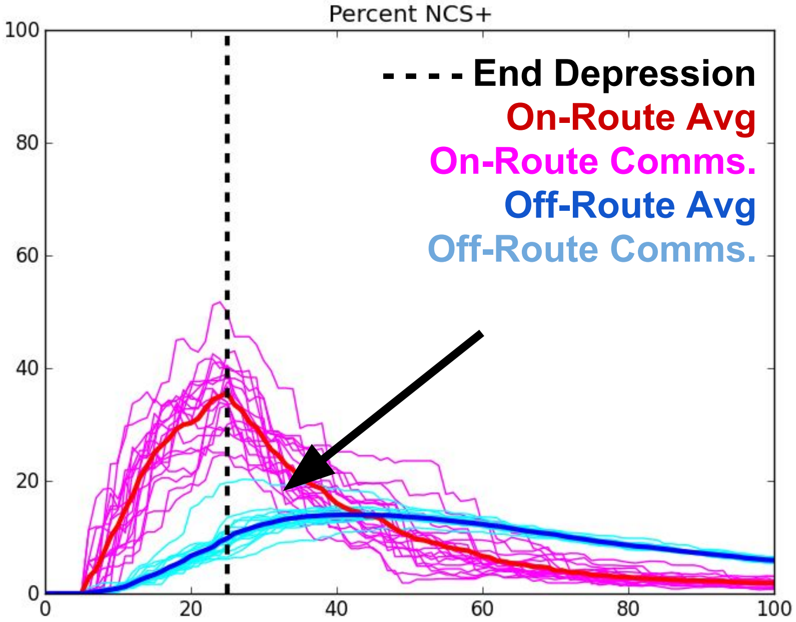
Results: Neutral Change

- A classic two-compartment pattern arises
- NCS peaks higher and earlier On-Route than Off-Route



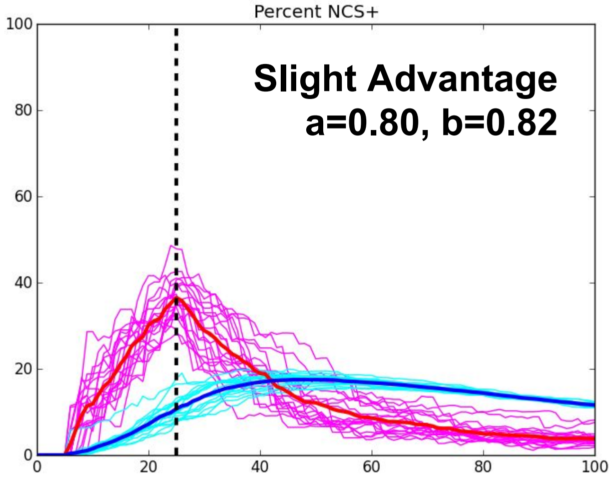
Results: Neutral Change

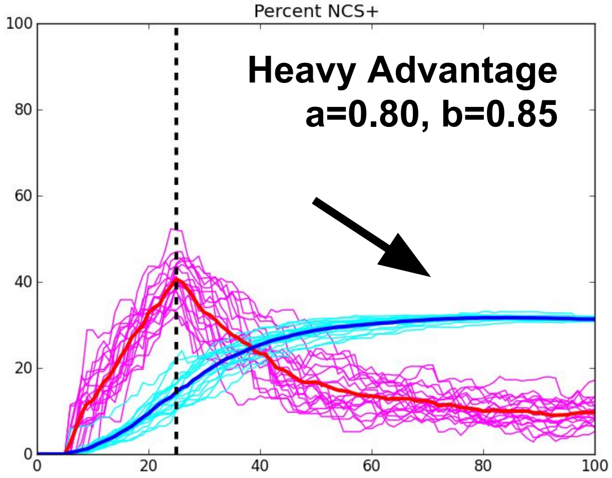
- A classic two-compartment pattern arises
- NCS peaks higher and earlier On-Route than Off-Route
- NCS continues to increase Off-Route even after On-Route population movements are reversed



Results: Advantaged Change

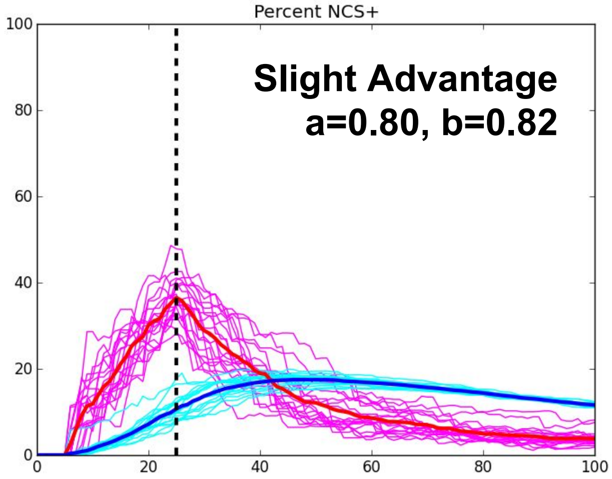
- **Advantaged change resists being "tamped down" Off-Route**
  - **NCS recedes given a slight advantage**
  - **NCS advances given a heavy advantage**

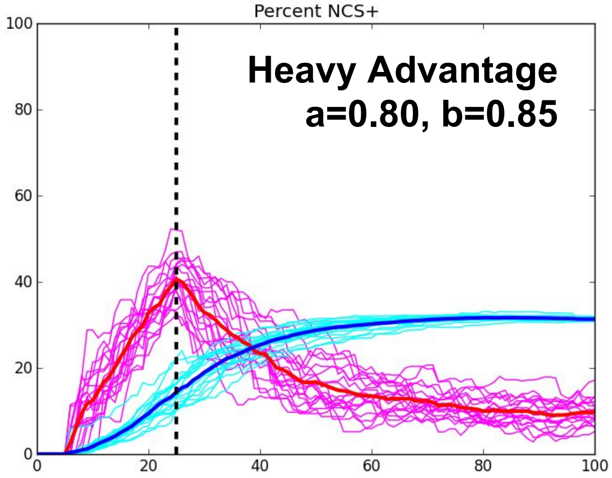




Results: Advantaged Change

- **Advantaged change resists being "tamped down" Off-Route**
  - **NCS recedes given a slight advantage**
  - **NCS advances given a heavy advantage**
- **Exists some threshold above which indirect action via On-Route is insufficient**





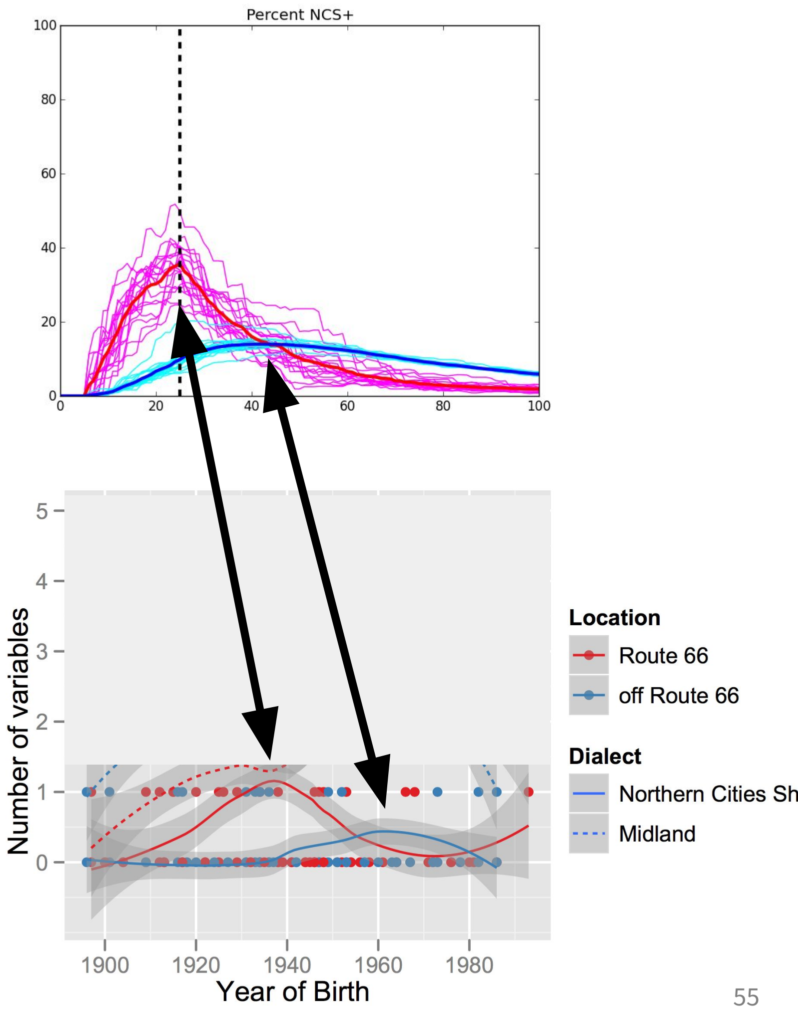
Analysis

Can Great Depression migrations account for the general path of change?

Analysis

Can Great Depression migrations account for the general path of change? **YES!**

- **Two-compartment pattern arises**

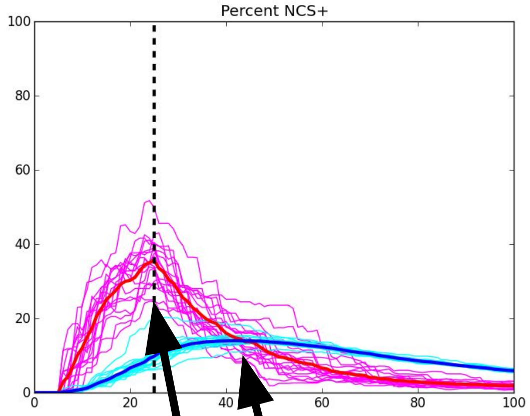


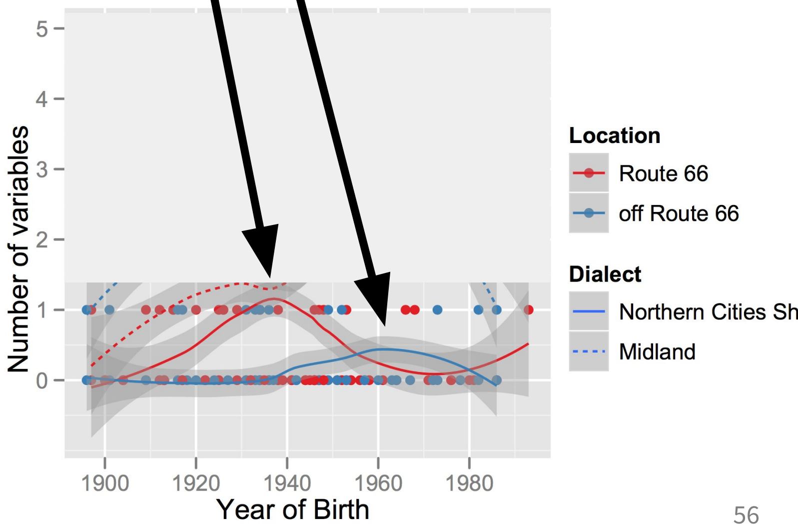
Analysis

Can Great Depression migrations account for the general path of change? **YES!**

- **Two-compartment pattern arises**

Was it the only factor?



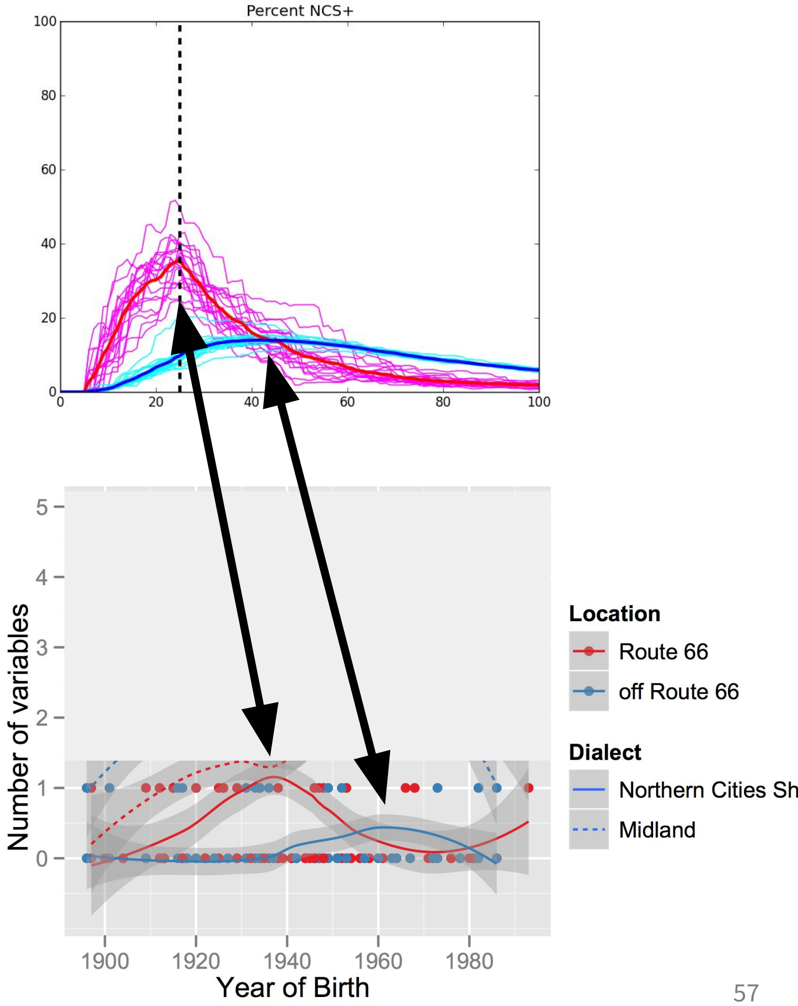


Analysis

Can Great Depression migrations account for the general path of change? **YES!**

- **Two-compartment pattern arises**

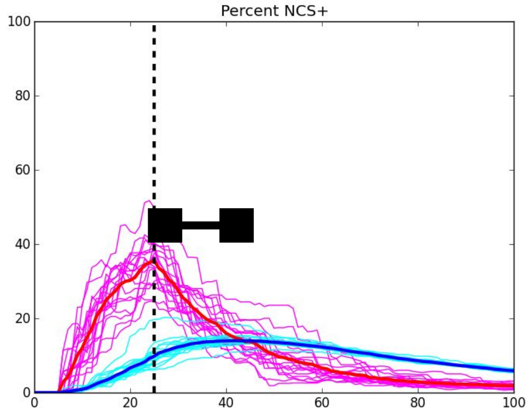
Was it the only factor? **NO...**

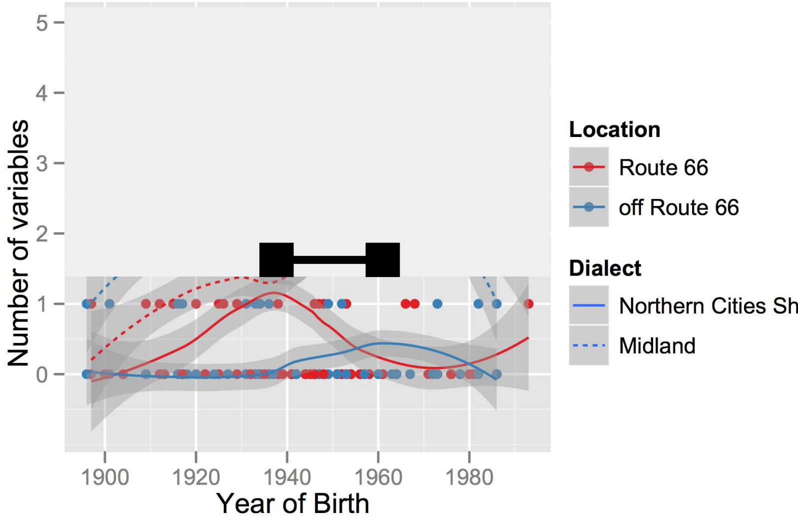


Discrepancies

Gap between the peaks

- **Due to our (overly) simple schematized network**





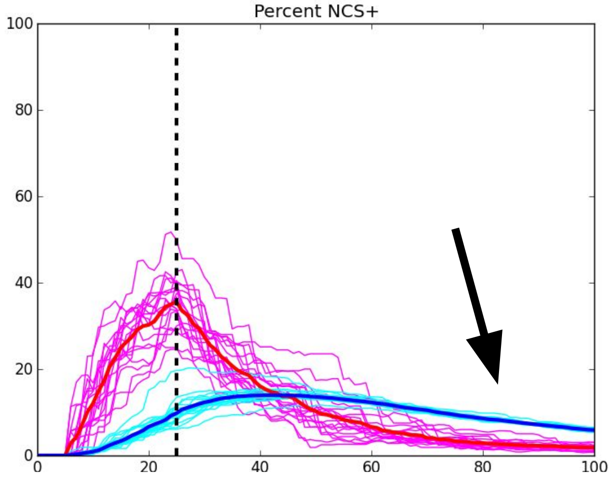
Discrepancies

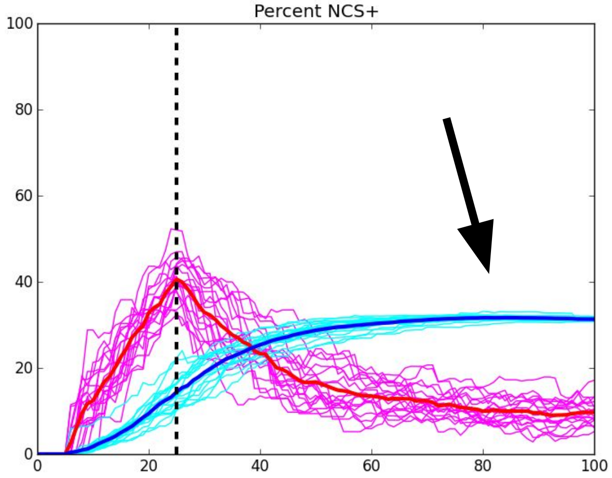
Gap between the peaks

- **Due to our (overly) simple schematized network**

Persistent NCS Off-Route

- **Could force it into the model…**
- **But pattern strongly suggestive of social factors**





Questions?

Code Available Here:

github.com/jkodner05/NetworksAndLangChange

Slides Available Here:

ling.upenn.edu/~jkodner

Special Thanks:
Chris Cerezo Falco
Charles Yang
ARO NDSEG

Extra slides: Math

Special Acknowledgement:

Christopher Cerezo Falco (UPenn)

Propagation

- A — *n* x *n* adjacency matrix
- H — *n* x *c* community-membership
- B — *c* x *g* distr. of grammars in comms
- P — *c* x *g* distr. of grammars in inputs
- α — jump parameter

Network Structure

- *n* x *n* adjacency matrix A
  - Value at a_{ji} indicates interaction from *j* to *i*
  - Must be column stochastic (columns sum to 1)
  - Undirected if for all *i*,*j*, $a_{ji} = a_{ij}$ (result row stochastic)

Propagation

Network Structure

- *n* x *n* adjacency matrix A
  - Value at a_{ji} indicates interaction from *j* to *i*
  - Must be column stochastic (columns sum to 1)
  - Undirected if for all *i*,*j*, $a_{ji} = a_{ij}$ (result row stochastic)
- *n* x *c* indicator matrix H
  - Values are 0 or 1
  - Identifies individuals as members of communities
  - H = I (n=c) if community membership is irrelevant

- A — *n* x *n* adjacency matrix
- H — *n* x *c* community-membership
- B — *c* x *g* distr. of grammars in comms
- P — *c* x *g* distr. of grammars in inputs
- α — jump parameter

- A n x n adjacency matrix
- H n x c community-membership
- B c x g distr. of grammars in comms
- P c x g distr. of grammars in inputs
- α jump parameter

Propagation

Distribution of Grammars

- c x g distribution of grammars (production) B_t
  - Value at b_{ij} indicates proportion of speaker j's outputs generated by grammar i
  - Must be column stochastic (columns sum to 1)
  - Indicator matrix if individuals only entertain one grammar

- **A** n x n adjacency matrix
- **H** n x c community-membership
- **B** c x g distr. of grammars in comms
- **P** c x g distr. of grammars in inputs
- **α** jump parameter

Propagation

Distribution of Grammars

- ***c*** **x** ***g*** **distribution of grammars (production)** B_t
  - **Value at** b_{ij} **indicates proportion of speaker j's outputs generated by grammar i**
  - **Must be column stochastic (columns sum to 1)**
  - **Indicator matrix if individuals only entertain one grammar**
- ***c*** **x** ***g*** **distribution of grammars (reception)** P_{t+1}
  - **Value at** p_{ij} **indicates proportion of speaker j's inputs generated by grammar i**
  - **Must be column stochastic (columns sum to 1)**
  - **Input to the acquisition algorithm**
  - **Calculated by Grammar Distribution Function**

Propagation

- **A** — *n* x *n* adjacency matrix
- **H** — *n* x *c* community-membership
- **B** — *c* x *g* distr. of grammars in comms
- **P** — *c* x *g* distr. of grammars in inputs
- **α** — jump parameter

Distribution of Grammars

- *c* x *g* distribution of grammars (production) B_t
  - Value at b_{ij} indicates proportion of speaker j's outputs generated by grammar i
  - Must be column stochastic (columns sum to 1)
  - Indicator matrix if individuals only entertain one grammar
- *c* x *g* distribution of grammars (reception) P_{t+1}
  - Value at p_{ij} indicates proportion of speaker j's inputs generated by grammar i
  - Must be column stochastic (columns sum to 1)
  - Input to the acquisition algorithm
  - Calculated by Grammar Distribution Function
- Scalar "jump" parameter α
  - Between 0 and 1

Intuition behind Calculation

- **Probability of interaction from *i* to *j* equals the probability of travelling from *i* to *j* along some path**

Intuition behind Calculation

- **Probability of interaction from *i* to *j* equals the probability of travelling from *i* to *j* along some path**
- **Edge weight is segment probability between adjacent *i* and *j***

Intuition behind Calculation

- **Probability of interaction from *i* to *j* equals the probability of travelling from *i* to *j* along some path**
- **Edge weight is segment probability between adjacent *i* and *j***
- **Probability of taking another jump decays according to geometric distribution**

Intuition behind Calculation

- **Probability of interaction from *i* to *j* equals the probability of travelling from *i* to *j* along some path**
- **Edge weight is segment probability between adjacent *i* and *j***
- **Probability of taking another jump decays according to geometric distribution**
- **Interaction likelihood decreases with social distance**

Grammar Distribution Function

$$\mathbf{P}_{t+1} = \mathbf{B}^\top \alpha (\mathbf{I} - (1-\alpha)\mathbf{A})^{-1} \mathbf{H}(\mathbf{H}^\top \mathbf{H})^{-1}$$

- **A** *n* x *n* adjacency matrix
- **α** jump parameter
- **H** *n* x *c* community-membership
- **B** *c* x *g* distr. of grammars in comms
- **P** *c* x *g* distr. of grammars in inputs

Grammar Distribution Function

$$\mathbf{P}_{t+1} = \mathbf{B}^\top \alpha (\mathbf{I} - (1-\alpha)\mathbf{A})^{-1} \mathbf{H}(\mathbf{H}^\top \mathbf{H})^{-1}$$

- **Single calculation for the entire population...FAST**

Grammar Distribution Function

$$\mathbf{P}_{t+1} = \mathbf{B}^\top \alpha \left(\mathbf{I} - (1-\alpha)\mathbf{A}\right)^{-1} \mathbf{H}(\mathbf{H}^\top \mathbf{H})^{-1}$$

- Single calculation for the entire population...FAST
- Requires an *n* x *n* matrix inversions...MEMORY INTENSIVE

Tracking Communities

- **If fine-grain detail is unnecessary, tracking community averages provides substantial computational speedup when *c* << *n***
- **If each community is internally uniform, *n* x *n* A admits a *c* x *c* equitable-partition A^π**
- **Yielding a more efficient but equivalent update formula for P**

$$
\begin{aligned}
\mathbf{A}^{\pi} &= (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{A} \mathbf{H} \\
\mathbf{P}_{t+1} &= \alpha \mathbf{B}^\top \mathbf{H} \left(\mathbf{I} - (1-\alpha)\mathbf{A}^{\pi}\right)^{-1} (\mathbf{H}^\top \mathbf{H})^{-1}
\end{aligned}
$$

Anecdotally, I can run *n* = 20,000 nets on my laptop with A^π about as fast as *n* = 2,000 net with A

Tracking Individuals

- **If $c = n$, then H is n x n, and the full descriptive detail of the model is available, H becomes the identity matrix, and the formula for P can be simplified**

$$\mathbf{P}_{t+1} = \mathbf{B}^\top \alpha (\mathbf{I} - (1-\alpha)\mathbf{A})^{-1}$$

Transmission

- **Dependent on the learning model**
- **Our implementation is modular, so many learning models can be slotted in**
  - **e.g., trigger-based learner (Gibson & Wexler 1994)**
  - **Variational learner (Yang 2000)**

Transmission

- Dependent on the learning model
- Our implementation is modular, so many learning models can be slotted in
  - e.g., trigger-based learner (Gibson & Wexler 1994)
  - Variational learner (Yang 2000)
- Let L be the distribution of grammars internalized by a learner who heard P
  - L is a matrix consisting of g vectors $l_1, l_2, \ldots l_g$
- Define *g* transition matrices $T_1, T_2, \ldots T_g$, one for each potential target grammar

$$\mathbf{l}_i = \text{dominant eigenvector of} \sum_{j=1}^{g} \mathbf{P}_{t+1;j,i} \mathbf{T}_j$$

Transmission and Grammatical Advantage

- **If L = P, learners internalize variants at the rate they hear them**
  - **This yields neutral change**
- **Otherwise, learners choose variants in a way that biases some over others**
  - **Some variants have an advantage over others**
  - **This yields S-curve change in perfectly mixed populations**

Transmission Example

- Let there be two languages L_1 and L_2, the extensions of g_1 and g_2, produced with probabilities P_1 and P_2.
- $a = P_1[L_1 \text{ union } L_2]$ $1 - a = P_1[L_1 \backslash L_2]$
- $b = P_2[L_1 \text{ union } L_2]$ $1 - b = P_2[L_2 \backslash L_1]$

Transmission Example

- **Let there be two languages L_1 and L_2, the extensions of g_1 and g_2, produced with probabilities P_1 and P_2.**
- **$a = P_1[L_1 \text{ union } L_2]$ $\quad 1 - a = P_1[L_1 \backslash L_2]$**
- **$b = P_2[L_1 \text{ union } L_2]$ $\quad 1 - b = P_2[L_2 \backslash L_1]$**
- **Let T_1 and T_2 be transition matrices assuming g_1 and g_2 are the target grammars respectively**
- **$T_1 = [1 \quad 0 \;;\; 1\text{-}a \quad a]$ $\quad T_2 = [b \quad 1\text{-}b \;;\; 0 \quad 1]$**

Transmission Example

$$T1 = \begin{bmatrix} 1 & 0 \\ 1-a & a \end{bmatrix}$$

$$T2 = \begin{bmatrix} b & 1-b \\ 0 & 1 \end{bmatrix}$$

- **If the target grammar is g1, then in the limit…**

Transmission Example

$$T1 = \begin{bmatrix} 1 & 0 \\ 1-a & a \end{bmatrix}$$

$$T2 = \begin{bmatrix} b & 1-b \\ 0 & 1 \end{bmatrix}$$

- **If the target grammar is g1, then in the limit...**
  - **Learners who initially hypothesize g1 will always remain in g1**

Transmission Example

$$T1 = \begin{bmatrix} 1 & 0 \\ 1-a & a \end{bmatrix}$$

$$T2 = \begin{bmatrix} b & 1-b \\ 0 & 1 \end{bmatrix}$$

- **If the target grammar is g1, then in the limit...**
  - **Learners who initially hypothesize g1 will always remain in g1**
  - **Learners who initially hypothesize g2 will remain at g2 with probability a**

Transmission Example

$$T1 = \begin{bmatrix} 1 & 0 \\ 1-a & a \end{bmatrix}$$

$$T2 = \begin{bmatrix} b & 1-b \\ 0 & 1 \end{bmatrix}$$

- **If the target grammar is g1, then in the limit...**
  - **Learners who initially hypothesize g1 will always remain in g1**
  - **Learners who initially hypothesize g2 will remain at g2 with probability a**
  - **Or switch to g1 with probability 1-a**