Analytics2011
CONFERENCE
SERIES

Dynamical Data Clustering

Carl Meyer, North Carolina State University

Cluster Analytics

- Classify objects from a set into clusters
- Objects in the same cluster are more similar to each other than to those in other clusters.
- Detect, reveal, and analyze hidden patterns

A Primary Exploratory Task

- Data and text mining
- Pattern recognition
- Image analysis
- Network analysis
- Information retrieval
- Bioinformatics

Fundamental Theorem Of Clustering

Nothing Works Always!

So Many Choices

- Hierarchical & agglomerative clustering
- k-means and its many variations and derivatives
- Nonnegative matrix factorization
- Spectral and subspace clustering
- Graph partitioning and min-cut techniques
- PDDP and PCA based partitioning algorithms
- Self organizing maps & neural network methods
- Gaussian mixture models & generalizations
- Nearest neighbor implementations
- Hard vs. fuzzy
- ··· more ··· and more ··· and more ···

Static Data vs. Dynamic Data

- Hidden patterns are more difficult define in static data

- But most clustering algorithms are built for the analysis of static data

Is It Relevant How The Fish Moves?

- By “fish intellegence?”
- Complicated system of PDE’s ?
- Strictly at random?

Is It Relevant How The Fish Moves?

By “fish intellegence?”

Complicated system of PDE’s ?

Strictly at random?

Doesn’t Matter

- As long as the spots on the fish stay attached to fish’s body
- Spots move together and relatively faster to slower movement of the background.

Dynamical Clustering

- Somehow impart “motion” to static data to reveal clusters moving by observing which data points move in concert relative to background noise.

Dynamical Clustering

- Somehow impart “motion” to static data to reveal clusters moving by observing which data points move in concert relative to background noise.

The Strategy

- Observe the evolution of a differentiated time-scale stochastic process imposed on the static data.

Dynamical Clustering

- Somehow impart “motion” to static data to reveal clusters moving by observing which data points move in concert relative to background noise.

The Strategy

- Observe the evolution of a differentiated time-scale stochastic process imposed on the static data.

How To Do It?

- Reverse Simon–Ando process

Simon–Ando Theory

- Herbert Simon (1916–2001)
- Carnegie Mellon University

— Nobel Prize in economics in 1978

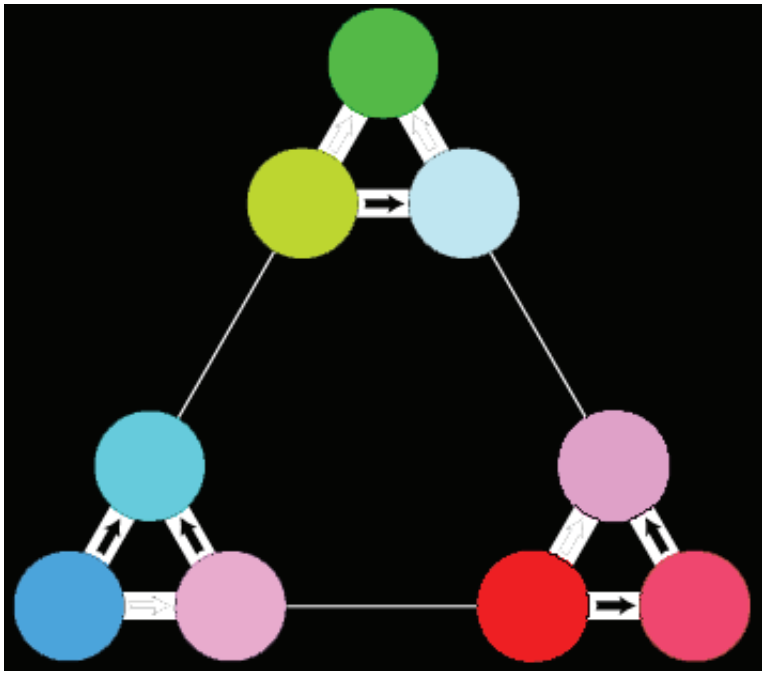
- Albert Ando (1930–2002)
- University of Pennsylvania

- "Aggregation of variables in dynamic systems," *Econometrica,* Vol. 29, No. 2 (Apr., 1961), pp. 111-138.

The Goal Of Simon–Ando

- Analyze long-term economic stability of a macro economy containing closely coupled micro economies by analyzing (or observing) the evolution of the micro economies for a short period of time.

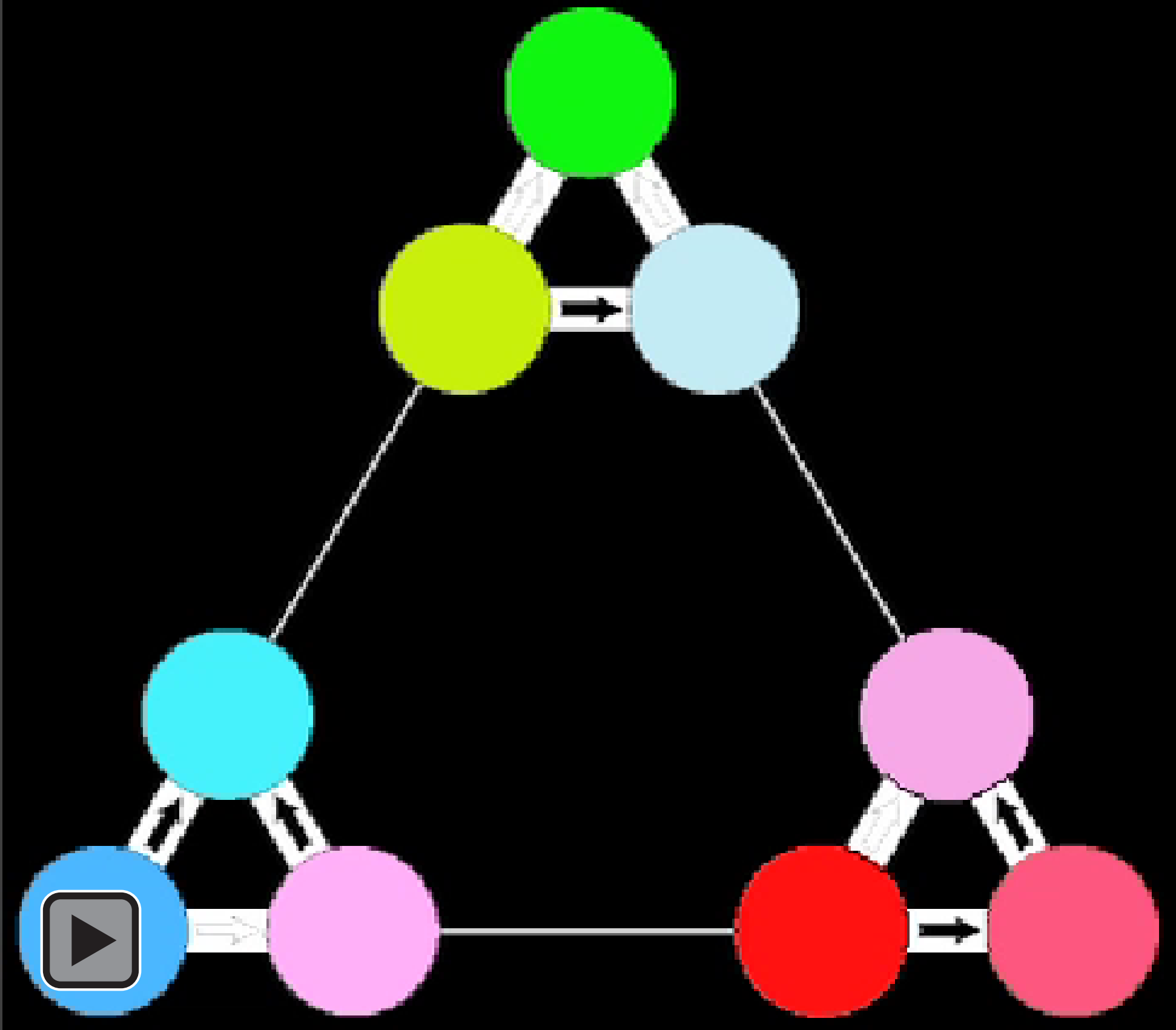
- Small example
  - Nine industries
  - Three closely coupled clusters



Cluster 1 = Manufacturing (steel, machine tools, heavy equipment)

Cluster 2 = Entertainment (movies, TV, books & magazines)

Cluster 3 = Beverage (sugar, water, packaging)



Simon–Ando Analysis

- Modeled by a Markov chain

Simon–Ando Analysis

Modeled by a Markov chain

- Individual industries are the states

Simon–Ando Analysis

Modeled by a Markov chain

Individual industries are the states

- Flow of capital between industries are the transitions

Simon–Ando Analysis

Modeled by a Markov chain

Individual industries are the states

Flow of capital between industries are the transitions

- Transition flows are row normalized (row sums = 1)
  - $\mathbf{P}_{n \times n}$ row stochastic transition matrix
  - Aperiodic Markov chain

Simon–Ando Analysis

Modeled by a Markov chain

Individual industries are the states

Flow of capital between industries are the transitions

Transition flows are row normalized (row sums = 1)

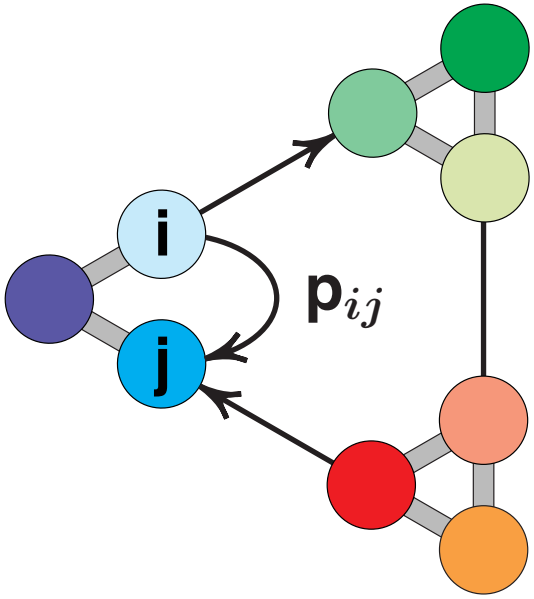
— $\mathbf{P}_{n\times n}$ row stochastic transition matrix

— Aperiodic Markov chain

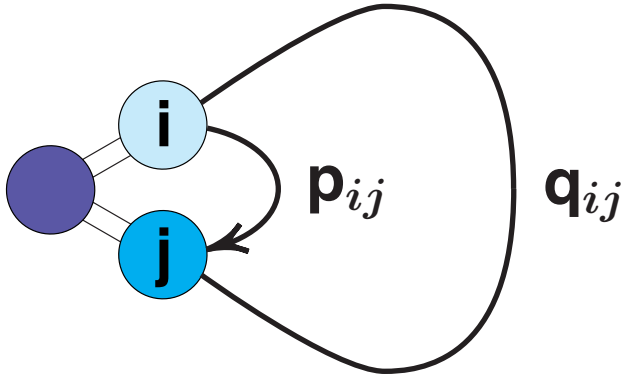
- k distinct micro economies (or clusters)
  - There is a permutation such that $\mathbf{P}$ is nearly uncoupled

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \cdots & \mathbf{P}_{1k} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \cdots & \mathbf{P}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{k1} & \mathbf{P}_{k2} & \cdots & \mathbf{P}_{kk} \end{bmatrix} \qquad \max_i ||\mathbf{P}_{i\star}||_\infty = \zeta << 1$$

Uncoupling By Censoring



$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \mathbf{P}_{13} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \mathbf{P}_{23} \\ \mathbf{P}_{31} & \mathbf{P}_{32} & \mathbf{P}_{33} \end{bmatrix}$$



(Uncensored) (Censored)

p_{ij} = P(*i* to *j* directly)

q_{ij} = P(reenter at *j* / leave from *i*)

$s_{ij} = p_{ij} + q_{ij}$ = Censored probability

- Censored Transition Matrices

$$\mathbf{S}_1 \equiv \mathbf{P}_{11} + \mathbf{Q}_1 \qquad \mathbf{S}_2 \equiv \mathbf{P}_{22} + \mathbf{Q}_2 \qquad \mathbf{S}_3 \equiv \mathbf{P}_{33} + \mathbf{Q}_3$$

Censoring For i=1

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \mathbf{P}_{13} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \mathbf{P}_{23} \\ \mathbf{P}_{31} & \mathbf{P}_{32} & \mathbf{P}_{33} \end{bmatrix}$$

- Censored probabilities

$$\mathbf{Q}_1 = [\mathbf{P}_{12} \quad \mathbf{P}_{13}] \begin{bmatrix} \mathbf{I} - \mathbf{P}_{22} & \mathbf{P}_{23} \\ \mathbf{P}_{32} & \mathbf{I} - \mathbf{P}_{33} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{P}_{21} \\ \mathbf{P}_{31} \end{bmatrix}$$

$$= \mathbf{P}_{1*} \left(\mathbf{I} - \widetilde{\mathbf{P}}_{11} \right)^{-1} \mathbf{P}_{*1}$$

Censoring For i=1

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \mathbf{P}_{13} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \mathbf{P}_{23} \\ \mathbf{P}_{31} & \mathbf{P}_{32} & \mathbf{P}_{33} \end{bmatrix}$$

- Censored probabilities

$$\mathbf{Q}_1 = [\mathbf{P}_{12} \quad \mathbf{P}_{13}] \begin{bmatrix} \mathbf{I} - \mathbf{P}_{22} & \mathbf{P}_{23} \\ \mathbf{P}_{32} & \mathbf{I} - \mathbf{P}_{33} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{P}_{21} \\ \mathbf{P}_{31} \end{bmatrix}$$

$$= \mathbf{P}_{1*} \left(\mathbf{I} - \widetilde{\mathbf{P}}_{11} \right)^{-1} \mathbf{P}_{*1}$$

- Censored transition matrix for censored chain

$$\mathbf{S}_1 = \mathbf{P}_{11} + \mathbf{Q}_1 = \mathbf{P}_{11} + \mathbf{P}_{1*} \left(\mathbf{I} - \widetilde{\mathbf{P}}_{11} \right)^{-1} \mathbf{P}_{*1}$$

Censoring For i=1

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \mathbf{P}_{13} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \mathbf{P}_{23} \\ \mathbf{P}_{31} & \mathbf{P}_{32} & \mathbf{P}_{33} \end{bmatrix}$$

- Censored probabilities

$$\mathbf{Q}_1 = [\mathbf{P}_{12} \quad \mathbf{P}_{13}] \begin{bmatrix} \mathbf{I} - \mathbf{P}_{22} & \mathbf{P}_{23} \\ \mathbf{P}_{32} & \mathbf{I} - \mathbf{P}_{33} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{P}_{21} \\ \mathbf{P}_{31} \end{bmatrix}$$

$$= \mathbf{P}_{1*} \left(\mathbf{I} - \widetilde{\mathbf{P}}_{11} \right)^{-1} \mathbf{P}_{*1}$$

- Censored transition matrix for censored chain

$$\mathbf{S}_1 = \mathbf{P}_{11} + \mathbf{Q}_1 = \mathbf{P}_{11} + \mathbf{P}_{1*} \left(\mathbf{I} - \widetilde{\mathbf{P}}_{11} \right)^{-1} \mathbf{P}_{*1}$$

- In General, $\mathbf{S}_i = \mathbf{P}_{ii} + \mathbf{P}_{i*} \left(\mathbf{I} - \widetilde{\mathbf{P}}_{ii} \right)^{-1} \mathbf{P}_{*i}$

- These are called ***stochastic complements***

Steady States

- $\boldsymbol{\sigma}_i$ = steady-state distribution of i^{th} censored chain

$\boldsymbol{\sigma}_i(0)$ = initial prob dist $\qquad$ $\boldsymbol{\sigma}_i(t) = \boldsymbol{\sigma}_i(t-1)\mathbf{S}_i$ = t-step dist

$$\boldsymbol{\sigma}_i = \lim_{t\to\infty} \boldsymbol{\sigma}_i(t) \qquad \boldsymbol{\sigma}_i = \boldsymbol{\sigma}_i\mathbf{S}_i$$

Steady States

$\boldsymbol{\sigma}_i$ = steady-state distribution of i^{th} censored chain

$\boldsymbol{\sigma}_i(0)$ = initial prob dist $\qquad$ $\boldsymbol{\sigma}_i(t) = \boldsymbol{\sigma}_i(t-1)\mathbf{S}_i$ = t-step dist

$\boldsymbol{\sigma}_i = \lim_{t\to\infty} \boldsymbol{\sigma}_i(t)$ $\qquad$ $\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_i\mathbf{S}_i$

- $\boldsymbol{\pi}$ = steady-state distribution of global chain

$\boldsymbol{\pi}(0)$ = initial prob dist $\qquad$ $\boldsymbol{\pi}(t) = \boldsymbol{\pi}(t-1)\mathbf{P}$ = t-step dist

$\boldsymbol{\pi} = \lim_{t\to\infty} \boldsymbol{\pi}(t)$ $\qquad$ $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$

Steady States

σ_i = steady-state distribution of i^{th} censored chain

$\sigma_i(0)$ = initial prob dist $\qquad$ $\sigma_i(t) = \sigma_i(t-1)\mathbf{S}_i$ = t-step dist

$\sigma_i = \lim_{t\to\infty} \sigma_i(t)$ $\qquad$ $\sigma_i = \sigma_i \mathbf{S}_i$

π = steady-state distribution of global chain

$\pi(0)$ = initial prob dist $\qquad$ $\pi(t) = \pi(t-1)\mathbf{P}$ = t-step dist

$\pi = \lim_{t\to\infty} \pi(t)$ $\qquad$ $\pi = \pi\mathbf{P}$

Coupling Theorem

- $\pi = (\,\xi_1\sigma_1 \;\; \xi_2\sigma_2 \;\; \xi_3\sigma_3\,)$, $\qquad$ ξ_1, ξ_2, ξ_3 are “coupling” constants

Equlibrium Phases

- **Short Run** $(a \leq t \leq b)$

$$\boldsymbol{\pi}(t) \approx \begin{pmatrix} \xi_1(t)\boldsymbol{\sigma_1} & \xi_2(t)\boldsymbol{\sigma_2} & \xi_3(t)\boldsymbol{\sigma_3} \end{pmatrix} \qquad \xi_i(t) = \|\boldsymbol{\pi}_i(0)\|_1 = \nu_i$$

Constants

Equlibrium Phases

Short Run $(a \leq t \leq b)$

$$\pi(t) \approx \begin{pmatrix} \xi_1(t)\boldsymbol{\sigma_1} & \xi_2(t)\boldsymbol{\sigma_2} & \xi_3(t)\boldsymbol{\sigma_3} \end{pmatrix} \qquad \xi_i(t) = \|\boldsymbol{\pi}_i(0)\|_1 = \nu_i$$

Constants

- **Middle Run** $(t > b)$

$$\boldsymbol{\pi}(t) \approx \begin{pmatrix} \xi_1(t)\boldsymbol{\sigma_1} & \xi_2(t)\boldsymbol{\sigma_2} & \xi_3(t)\boldsymbol{\sigma_3} \end{pmatrix} \qquad \xi_i(t) \text{ Varies With Time}$$

Equlibrium Phases

Short Run $(a \leq t \leq b)$

$$\boldsymbol{\pi}(t) \approx \begin{pmatrix} \xi_1(t)\boldsymbol{\sigma}_1 & \xi_2(t)\boldsymbol{\sigma}_2 & \xi_3(t)\boldsymbol{\sigma}_3 \end{pmatrix} \qquad \xi_i(t) = \|\boldsymbol{\pi}_i(0)\|_1 = \nu_i$$

Constants

Middle Run $(t > b)$

$$\boldsymbol{\pi}(t) \approx \begin{pmatrix} \xi_1(t)\boldsymbol{\sigma}_1 & \xi_2(t)\boldsymbol{\sigma}_2 & \xi_3(t)\boldsymbol{\sigma}_3 \end{pmatrix} \qquad \xi_i(t) \text{ Varies With Time}$$

For example, consider states i and j in cluster #1

$$\frac{\boldsymbol{\pi}_i(t)}{\boldsymbol{\pi}_j(t)} \approx \frac{\xi_1(t)[\boldsymbol{\sigma}_1]_i}{\xi_1(t)[\boldsymbol{\sigma}_1]_j} = \frac{[\boldsymbol{\sigma}_1]_i}{[\boldsymbol{\sigma}_1]_j} = \text{ a constant} \qquad (t > a)$$

Equlibrium Phases

Short Run $(a \leq t \leq b)$

$$\boldsymbol{\pi}(t) \approx \begin{pmatrix} \xi_1(t)\boldsymbol{\sigma}_1 & \xi_2(t)\boldsymbol{\sigma}_2 & \xi_3(t)\boldsymbol{\sigma}_3 \end{pmatrix} \qquad \xi_i(t) = \|\boldsymbol{\pi}_i(0)\|_1 = \nu_i$$

Constants

Middle Run $(t > b)$

$$\boldsymbol{\pi}(t) \approx \begin{pmatrix} \xi_1(t)\boldsymbol{\sigma}_1 & \xi_2(t)\boldsymbol{\sigma}_2 & \xi_3(t)\boldsymbol{\sigma}_3 \end{pmatrix} \qquad \xi_i(t) \text{ Varies With Time}$$

For example, consider states i and j in cluster #1

$$\frac{\boldsymbol{\pi}_i(t)}{\boldsymbol{\pi}_j(t)} \approx \frac{\xi_1(t)[\boldsymbol{\sigma}_1]_i}{\xi_1(t)[\boldsymbol{\sigma}_1]_j} = \frac{[\boldsymbol{\sigma}_1]_i}{[\boldsymbol{\sigma}_1]_j} = \text{ a constant} \qquad (t > a)$$

- **Long Run** $\xi_i(t) \longrightarrow \xi_i$ (Constant) as $t \longrightarrow \infty$

$$\boldsymbol{\pi}(t) \longrightarrow \boldsymbol{\pi}(\infty) = \begin{pmatrix} \xi_1\boldsymbol{\sigma}_1 & \xi_2\boldsymbol{\sigma}_2 & \xi_3\boldsymbol{\sigma}_3 \end{pmatrix} = \boldsymbol{\pi} \qquad \text{(Global Equlibrium)}$$

Simon–Ando Conclusions

- Short-run behavior reveals long-run behavior

Simon–Ando Conclusions

Short-run behavior reveals long-run behavior

- Long-run equlibrium in a macro economy containing clusters of micro economies is determined by the short-run evolution of the micros.

Simon–Ando Conclusions

Short-run behavior reveals long-run behavior

Long-run equlibrium in a macro economy containing clusters of micro economies is determined by the short-run evolution of the micros.

- Longer-term economic predictions can be made from shorter-term observations.

Reverse Simon–Ando

- Long-run behavior reveals short-run behavior

$$\boldsymbol{\pi}(\infty) = \boldsymbol{\pi} = \begin{pmatrix} \boldsymbol{\pi}_1 & \boldsymbol{\pi}_2 & \boldsymbol{\pi}_3 \end{pmatrix} \implies \boldsymbol{\sigma}_i = \frac{\boldsymbol{\pi}_i}{\|\boldsymbol{\pi}_i\|_1}$$

Reverse Simon–Ando

Long-run behavior reveals short-run behavior

$$\pi(\infty) = \pi = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \end{pmatrix} \Longrightarrow \sigma_i = \frac{\pi_i}{\|\pi_i\|_1}$$

$$\pi(a \leq t \leq b) \approx \begin{pmatrix} \nu_1\sigma_1 & \nu_2\sigma_2 & \nu_3\sigma_3 \end{pmatrix} \quad \text{where} \quad \nu_i = \|\pi_i(0)\|_1$$

(Short-run stabilization)

Reverse Simon–Ando

Long-run behavior reveals short-run behavior

$$\boldsymbol{\pi}(\infty) = \boldsymbol{\pi} = (\,\boldsymbol{\pi}_1 \quad \boldsymbol{\pi}_2 \quad \boldsymbol{\pi}_3\,) \implies \boldsymbol{\sigma}_i = \frac{\boldsymbol{\pi}_i}{\|\boldsymbol{\pi}_i\|_1}$$

$$\boldsymbol{\pi}(a \le t \le b) \approx (\,\nu_1\boldsymbol{\sigma}_1 \quad \nu_2\boldsymbol{\sigma}_2 \quad \nu_3\boldsymbol{\sigma}_3\,) \quad \text{where} \quad \nu_i = \|\boldsymbol{\pi}_i(0)\|_1$$

(Short-run stabilization)

- And middle-run behavior

$$\boldsymbol{\pi}(t > b) \approx (\,\xi_1(t)\boldsymbol{\sigma}_1 \quad \xi_2(t)\boldsymbol{\sigma}_2 \quad \xi_3(t)\boldsymbol{\sigma}_3\,) \quad \text{where} \quad \xi_i(t) \to \|\boldsymbol{\pi}_i\|_1$$

Revealing Clusters

- Suppose that $\pi(\infty)$ is uniform

$$\pi(\infty) = \frac{1}{n}\left(\overbrace{1,1,\ldots,1,}^{\mathbf{n}_1}\ \overbrace{1,1,\ldots,1,}^{\mathbf{n}_2}\ \overbrace{1,1,\ldots,1}^{\mathbf{n}_3}\right) = \begin{pmatrix} \boldsymbol{\pi_1} & \boldsymbol{\pi_2} & \boldsymbol{\pi_3} \end{pmatrix} \qquad \mathbf{n}_i = \#\ \textbf{states in cluster}\ i$$

Revealing Clusters

Suppose that $\pi(\infty)$ is uniform

$$\pi(\infty) = \frac{1}{n}\left(\overbrace{1,1,\ldots,1}^{n_1},\ \overbrace{1,1,\ldots,1}^{n_2},\ \overbrace{1,1,\ldots,1}^{n_3}\right) = \left(\pi_1 \quad \pi_2 \quad \pi_3\right) \qquad n_i = \#\ \text{states in cluster}\ i$$

- $\boldsymbol{\sigma}_i = \boldsymbol{\pi}_i / \left\|\boldsymbol{\pi}_i\right\|_1 = \dfrac{1}{n_i}\left(\overbrace{1, 1, \ldots, 1}^{n_i}\right)$

Revealing Clusters

Suppose that $\pi(\infty)$ is uniform

$$\pi(\infty) = \frac{1}{n}\,(\overbrace{1,1,\ldots,1}^{\mathbf{n}_1},\ \overbrace{1,1,\ldots,1}^{\mathbf{n}_2},\ \overbrace{1,1,\ldots,1}^{\mathbf{n}_3}) = (\,\boldsymbol{\pi_1}\quad \boldsymbol{\pi_2}\quad \boldsymbol{\pi_3}\,) \qquad \mathbf{n}_i = \#\ \text{states in cluster}\ i$$

$$\boldsymbol{\sigma}_i = \boldsymbol{\pi}_i / \left\|\boldsymbol{\pi}_i\right\|_1 = \frac{1}{n_i}\,(\overbrace{1,1,\ldots,1}^{\mathbf{n}_i})$$

$\alpha_i = \nu_i / n_i$

- $\boldsymbol{\pi}(a \le t \le b) \approx (\,\nu_1\boldsymbol{\sigma_1}\quad \nu_2\boldsymbol{\sigma_2}\quad \nu_3\boldsymbol{\sigma_3}\,) = (\,\alpha_1 \cdots \alpha_1 \,|\, \alpha_2 \cdots \alpha_2 \,|\, \alpha_3 \cdots \alpha_3\,)$

Revealing Clusters

Suppose that $\pi(\infty)$ is uniform

$$\pi(\infty) = \frac{1}{n}\,(\overbrace{1,1,\ldots,1}^{\mathbf{n_1}},\ \overbrace{1,1,\ldots,1}^{\mathbf{n_2}},\ \overbrace{1,1,\ldots,1}^{\mathbf{n_3}}) = (\,\boldsymbol{\pi_1}\quad\boldsymbol{\pi_2}\quad\boldsymbol{\pi_3}\,) \qquad \mathbf{n}_i = \#\ \text{states in cluster}\ i$$

$$\boldsymbol{\sigma}_i = \boldsymbol{\pi}_i / \left\|\boldsymbol{\pi}_i\right\|_1 = \frac{1}{n_i}\,(\overbrace{1,1,\ldots,1}^{\mathbf{n}_i})$$

$$\pi(a \le t \le b) \approx (\,\nu_1\boldsymbol{\sigma}_1 \quad \nu_2\boldsymbol{\sigma}_2 \quad \nu_3\boldsymbol{\sigma}_3\,) = (\,\alpha_1\cdots\alpha_1 \,|\, \alpha_2\cdots\alpha_2 \,|\, \alpha_3\cdots\alpha_3\,) \qquad \alpha_i = \nu_i/n_i$$

- $\boldsymbol{\pi}(t > b) \approx (\,\xi_1(t)\boldsymbol{\sigma}_1 \quad \xi_2(t)\boldsymbol{\sigma}_2 \quad \xi_3(t)\boldsymbol{\sigma}_3\,)$

  $= (\,\beta_1(t)\cdots\beta_1(t) \,|\, \beta_2(t)\cdots\beta_2(t) \,|\, \beta_3(t)\cdots\beta_3(t)\,)$ $\qquad \beta_i(t) \to 1/n$

Revealing Clusters

Suppose that $\pi(\infty)$ is uniform

$$\pi(\infty) = \frac{1}{n}\,(\overbrace{1,1,\ldots,1}^{\mathbf{n_1}},\ \overbrace{1,1,\ldots,1}^{\mathbf{n_2}},\ \overbrace{1,1,\ldots,1}^{\mathbf{n_3}}) = (\,\boldsymbol{\pi_1}\quad \boldsymbol{\pi_2}\quad \boldsymbol{\pi_3}\,) \qquad \mathbf{n}_i = \#\ \text{states in cluster}\ i$$

$$\boldsymbol{\sigma}_i = \boldsymbol{\pi}_i / \left\|\boldsymbol{\pi}_i\right\|_1 = \frac{1}{n_i}\,(\overbrace{1,1,\ldots,1}^{\mathbf{n}_i})$$

$$\pi(a \le t \le b) \approx (\,\nu_1\boldsymbol{\sigma}_1 \quad \nu_2\boldsymbol{\sigma}_2 \quad \nu_3\boldsymbol{\sigma}_3\,) = (\,\alpha_1\cdots\alpha_1 \,|\, \alpha_2\cdots\alpha_2 \,|\, \alpha_3\cdots\alpha_3\,) \qquad \alpha_i = \nu_i/n_i$$

$$\pi(t > b) \approx (\,\xi_1(t)\boldsymbol{\sigma}_1 \quad \xi_2(t)\boldsymbol{\sigma}_2 \quad \xi_3(t)\boldsymbol{\sigma}_3\,)$$

$$= (\,\beta_1(t)\cdots\beta_1(t) \,|\, \beta_2(t)\cdots\beta_2(t) \,|\, \beta_3(t)\cdots\beta_3(t)\,) \qquad \beta_i(t) \to 1/n$$

- ***Nearly equal entries in $\pi(t > a)$ belong to the same cluster***

Question

- How Do We Force $\pi(\infty)$ To Be Uniform?

Question

How Do We Force $\pi(\infty)$ To Be Uniform?

Answer

- Simply force **P** to be doubly stochastic
  - i.e., force all row sums and all column sums = 1

Question

How Do We Force $\pi(\infty)$ To Be Uniform?

Answer

Simply force **P** to be doubly stochastic

— i.e., force all row sums and all column sums = 1

Sinkhorn–Knopp Procedure

- Scale rows

Question

How Do We Force $\pi(\infty)$ To Be Uniform?

Answer

Simply force **P** to be doubly stochastic

— i.e., force all row sums and all column sums = 1

Sinkhorn–Knopp Procedure

- Scale rows $\longrightarrow$ scale columns

Question

How Do We Force $\pi(\infty)$ To Be Uniform?

Answer

Simply force **P** to be doubly stochastic

— i.e., force all row sums and all column sums = 1

Sinkhorn–Knopp Procedure

- Scale rows $\longrightarrow$ scale columns $\longrightarrow$ scale rows

Question

How Do We Force $\pi(\infty)$ To Be Uniform?

Answer

Simply force **P** to be doubly stochastic

— i.e., force all row sums and all column sums = 1

Sinkhorn–Knopp Procedure

- Scale rows $\longrightarrow$ scale columns $\longrightarrow$ scale rows $\longrightarrow$ scale columns

Question

How Do We Force $\pi(\infty)$ To Be Uniform?

Answer

Simply force **P** to be doubly stochastic

— i.e., force all row sums and all column sums = 1

Sinkhorn–Knopp Procedure

- Scale rows $\longrightarrow$ scale columns $\longrightarrow$ scale rows $\longrightarrow$ scale columns $\longrightarrow$ etc.

Question

How Do We Force $\pi(\infty)$ To Be Uniform?

Answer

Simply force **P** to be doubly stochastic

— i.e., force all row sums and all column sums = 1

Sinkhorn–Knopp Procedure

- Scale rows $\longrightarrow$ scale columns $\longrightarrow$ scale rows $\longrightarrow$ scale columns $\longrightarrow$ etc.
- Converges (usually) — if not, it can be forced

Question

How Do We Force $\pi(\infty)$ To Be Uniform?

Answer

Simply force **P** to be doubly stochastic

— i.e., force all row sums and all column sums = 1

Sinkhorn–Knopp Procedure

- Scale rows $\longrightarrow$ scale columns $\longrightarrow$ scale rows $\longrightarrow$ scale columns $\longrightarrow$ etc.

  Converges (usually) — if not, it can be forced

- Sinkhorn–Knopp preserves cluster integrity

Question

How Do We Force $\pi(\infty)$ To Be Uniform?

Answer

Simply force **P** to be doubly stochastic

— i.e., force all row sums and all column sums = 1

Sinkhorn–Knopp Procedure

- Scale rows $\longrightarrow$ scale columns $\longrightarrow$ scale rows $\longrightarrow$ scale columns $\longrightarrow$ etc.

Converges (usually) — if not, it can be forced

Sinkhorn–Knopp preserves cluster integrity

- Sinkhorn–Knopp preserves symmetry

Putting Things Together

Raw Data

- Step 1. $\mathbf{A}_{mn} = [\,\mathbf{a}_1 \,|\, \mathbf{a}_2 \,|\, \cdots \,|\, \mathbf{a}_n\,] \;\rightarrow \mathbf{C}_{nn}$ (A Similarity Matrix)

  $c_{ij} = d(\mathbf{a}_i, \mathbf{a}_j)$ $\qquad$ $d(\star, \star)$ = similarity metric of your choice

Putting Things Together

Raw Data

- Step 1. $\mathbf{A}_{mn} = [\,\mathbf{a}_1 \,|\, \mathbf{a}_2 \,|\, \cdots \,|\, \mathbf{a}_n\,] \;\rightarrow \mathbf{C}_{nn}$ (A Similarity Matrix)

$c_{ij} = d(\mathbf{a}_i, \mathbf{a}_j)$ $\qquad$ $d(\star, \star)$ = similarity metric of your choice

- Euclidean distance
- Minkowski norms
- Correlation
- Angular distance
- Gaussian metrics
- Hamming distance or variations
- Gabriel graph
- Delaunay triangulation
- Mean first passage time
- Ensemble (consensus) metrics
- etc

Putting Things Together

Raw Data

- Step 1. $\mathbf{A}_{mn} = [\,\mathbf{a}_1 \,|\, \mathbf{a}_2 \,|\, \cdots \,|\, \mathbf{a}_n\,] \;\rightarrow \mathbf{C}_{nn}$ (A Similarity Matrix)

$c_{ij} = d(\mathbf{a}_i, \mathbf{a}_j)$ $\qquad$ $d(\star, \star)$ = similarity metric of your choice

- Euclidean distance
- Minkowski norms
- Correlation
- Angular distance
- Gaussian metrics
- Hamming distance or variations
- Gabriel graph
- Delaunay triangulation
- Mean first passage time
- Ensemble (consensus) metrics
- etc

— Applications usually dictates choice

— e.g., text vs. numeric

— $\mathbf{C} = \mathbf{C}^T$ (symmetric matrix—usually)

- Step 2. **Sinkhorn–Knopp Procedure**
  - $\mathbf{C}_{nn} \longrightarrow \mathbf{P}_{nn}$ (by successive row & column scaling)
  - Similarity matrix (Symmetric) $\longrightarrow$ doubly stochastic (Symmetric)

Step 2. **Sinkhorn–Knopp Procedure**

— $\mathbf{C}_{nn} \longrightarrow \mathbf{P}_{nn}$ (by successive row & column scaling)

— Similarity matrix (SYMMETRIC) $\longrightarrow$ doubly stochastic (SYMMETRIC)

- Step 3. **Initialize Markov Chain**
  - Pick $\pi(0)$ to be significantly different than uniform
  - To see where state (or data point) $\#i$ clusters, pick

$$\pi(0) = \mathbf{e}_i = (0, 0, \ldots, \underset{\substack{\uparrow \\ i}}{1}, 0, \ldots, 0)$$

Step 2. **Sinkhorn–Knopp Procedure**

— $\mathbf{C}_{nn} \longrightarrow \mathbf{P}_{nn}$ (by successive row & column scaling)

— Similarity matrix (Symmetric) $\longrightarrow$ doubly stochastic (Symmetric)

Step 3. **Initialize Markov Chain**

— Pick $\pi(0)$ to be significantly different than uniform

— To see where state (or data point) $\#i$ clusters, pick

$$\pi(0) = \mathbf{e}_i = (0, 0, \ldots, \underset{\substack{\uparrow \\ i}}{1}, 0, \ldots, 0)$$

- Step 4. **Observe Markov Chain For A Few Steps**

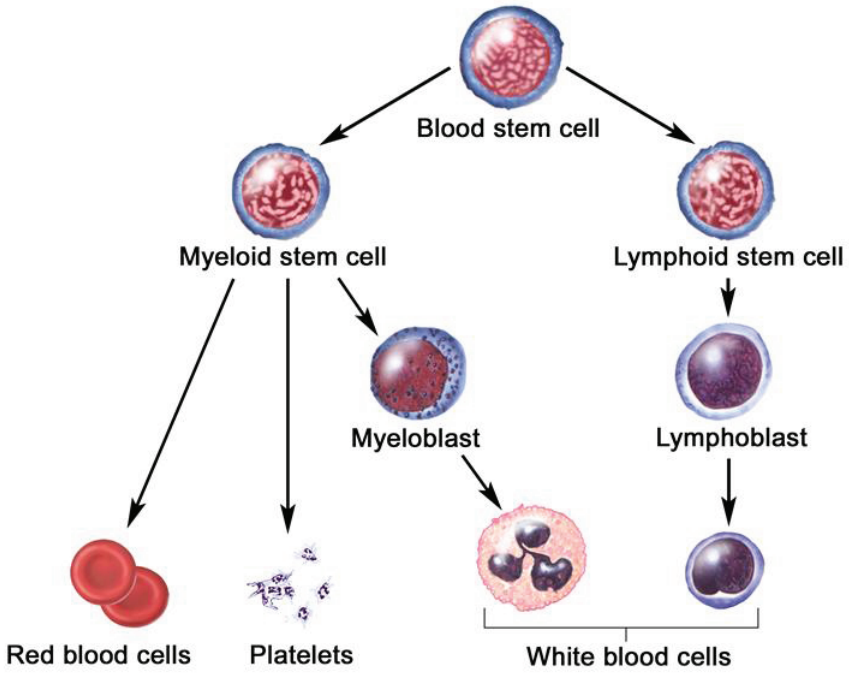
  — $\pi(t+1) = \pi(t)\mathbf{P}$ $\quad t = 1, 2, \ldots, a$ (until short-run stabilization)

  — Order entries in $\pi(t > a)$ to identify gaps

  — Nearly equal entries belong to the same cluster

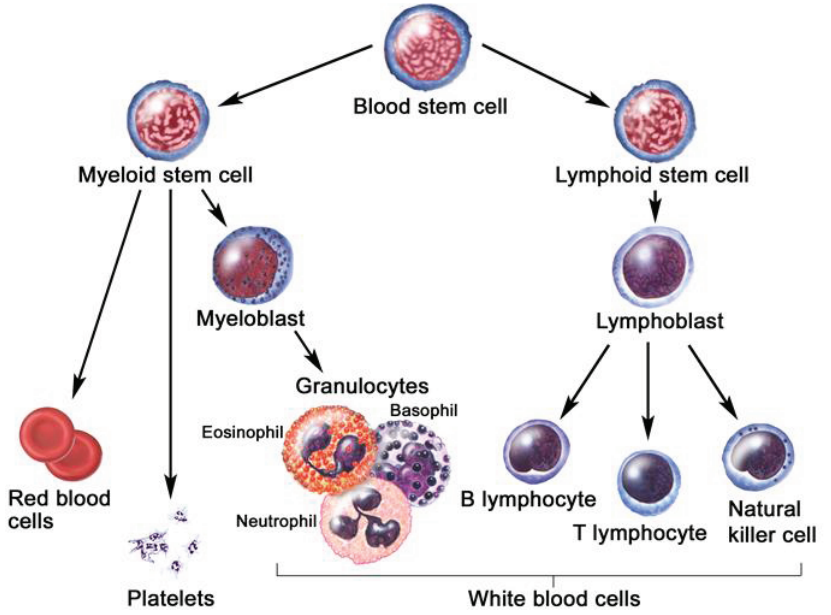
Leukemia Exmple

- AML — Acute myeloid leukemia (Most common in adults)



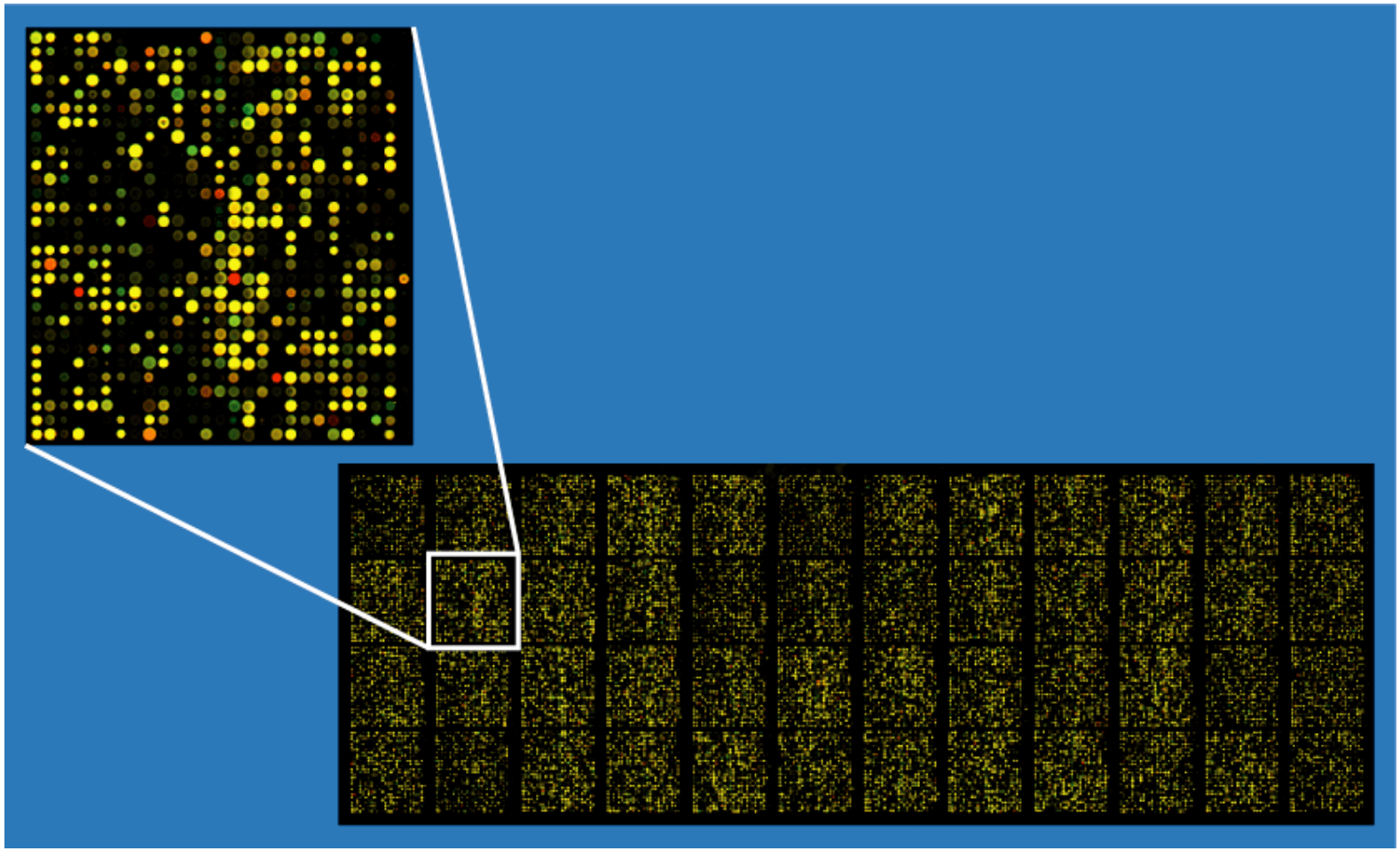
In AML, myeloid stem cells develop into immature abnormal white blood cells, myeloblasts, that don't become healthy white blood cells

- ALL — Acute lymphoblastic leukemia (Most common in children)



In ALL, too many stem cells develop into lymphoblasts that don't mature into lymphocytes, the white blood cells required to fight infections.

DNA Microarray



Broad Institute (MIT/Harvard)

- "Molecular classifcation of cancer: class discovery and class prediction by gene expression monitoring," T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfeld, and E. S. Lander, *Science,* October 1999.

- "Metagenes and molecular pattern discovery using matrix factorization," J. P. Brunet, P. Tamayo, T. Golub, J. Mesirov, *Proceedings of the National Academy of Sciences,* March, 2004.

- 38 cancer patients — gene expression data from bone marrow samples — 5000 genes

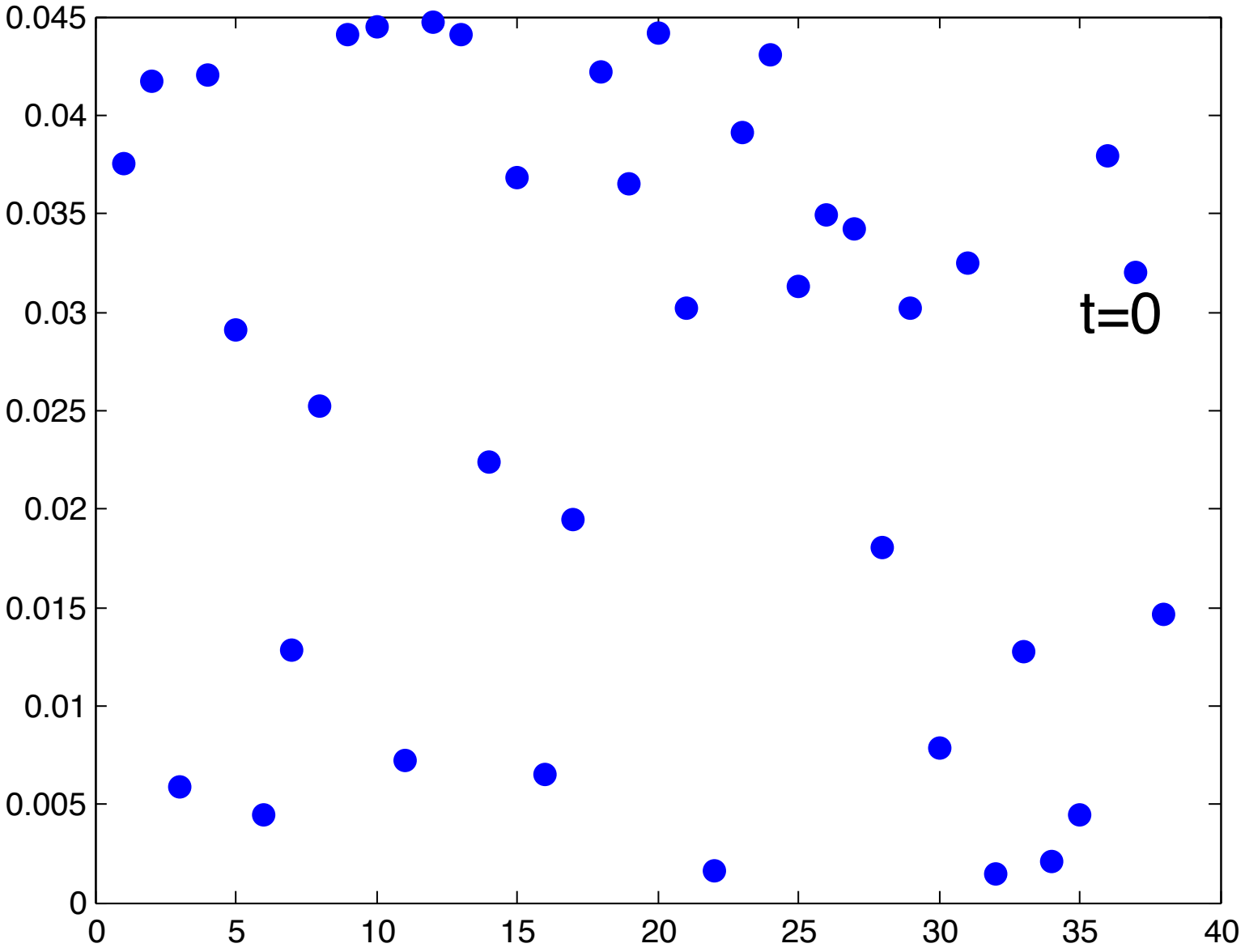
- Patient Diagonsis: #1–19 = ALL(B), #20–27 = ALL(T), #28–38 = ALM

- Good test case since clusters are known

- Similarity matrix was build by a consensus method

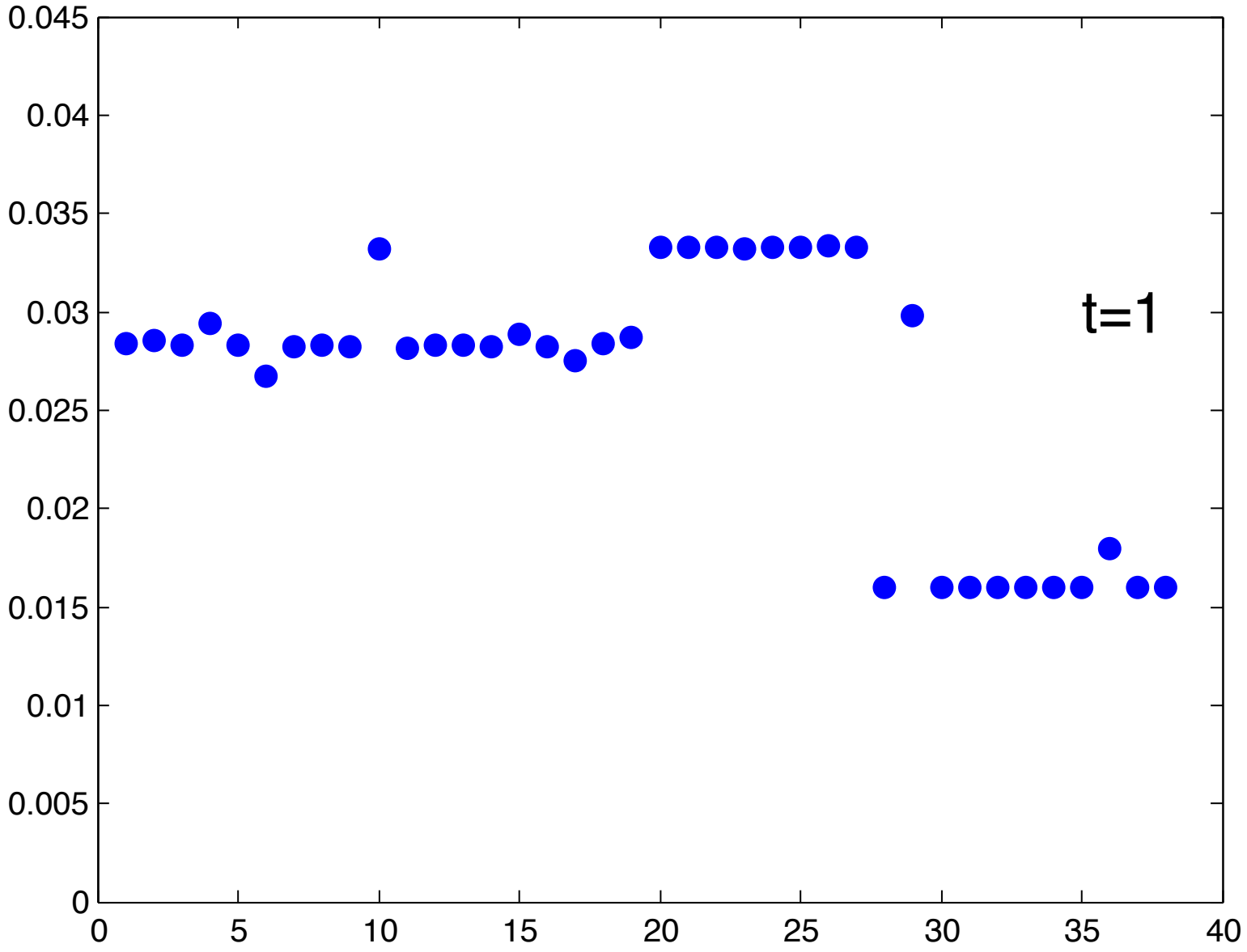
The Iterations

(Chuck Wessell, NC State)



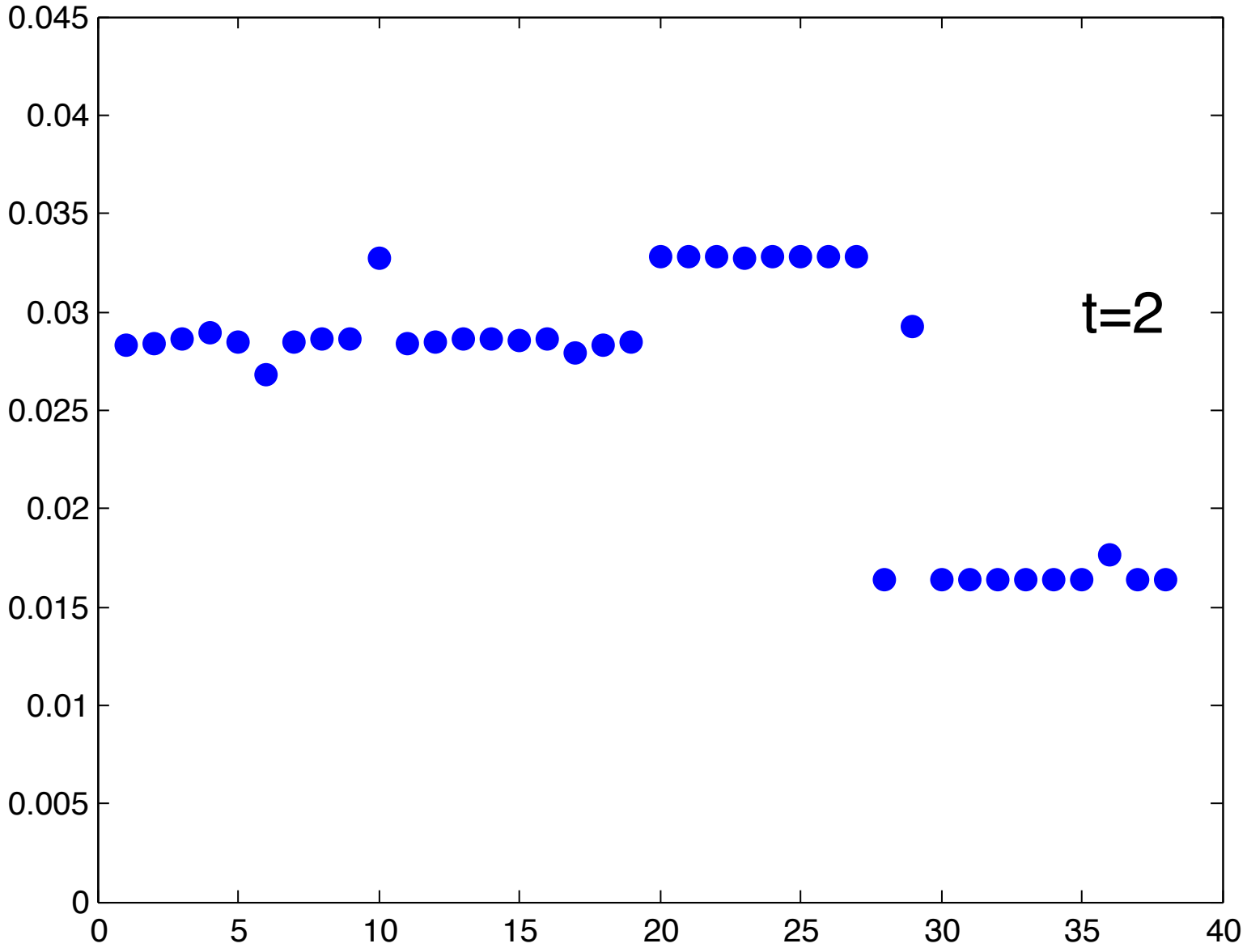
The Iterations

(Chuck Wessell, NC State)



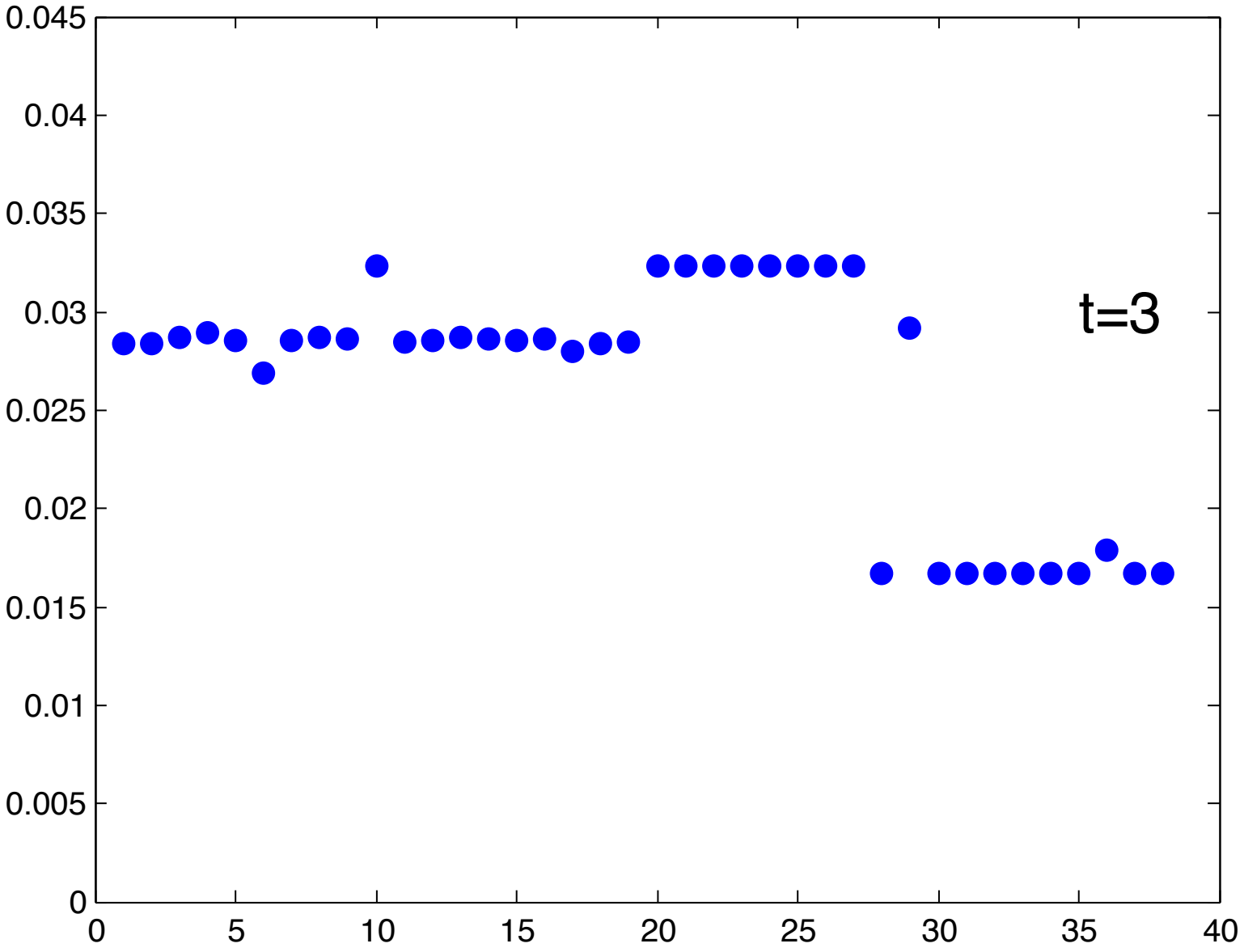
The Iterations

(Chuck Wessell, NC State)



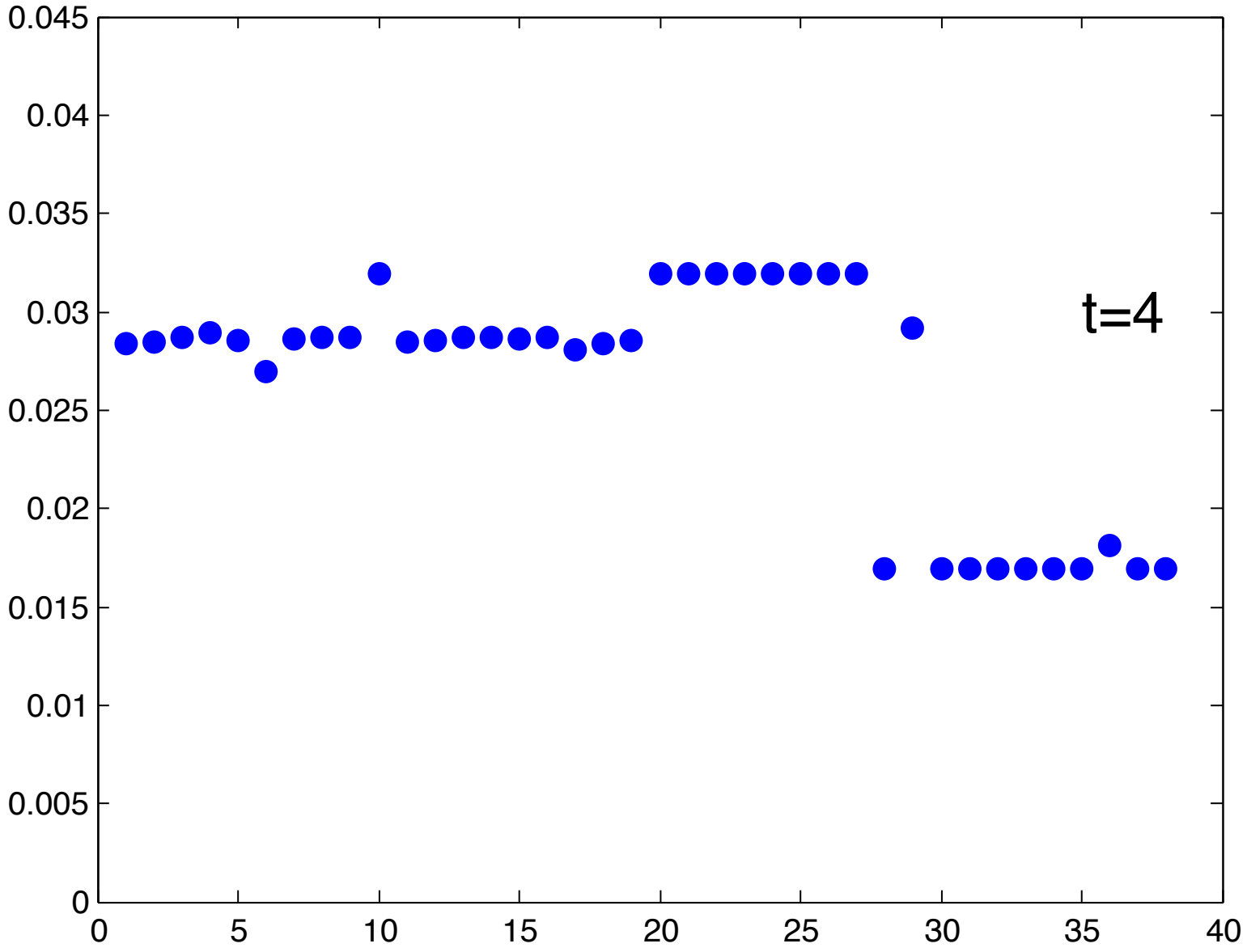
The Iterations

(Chuck Wessell, NC State)



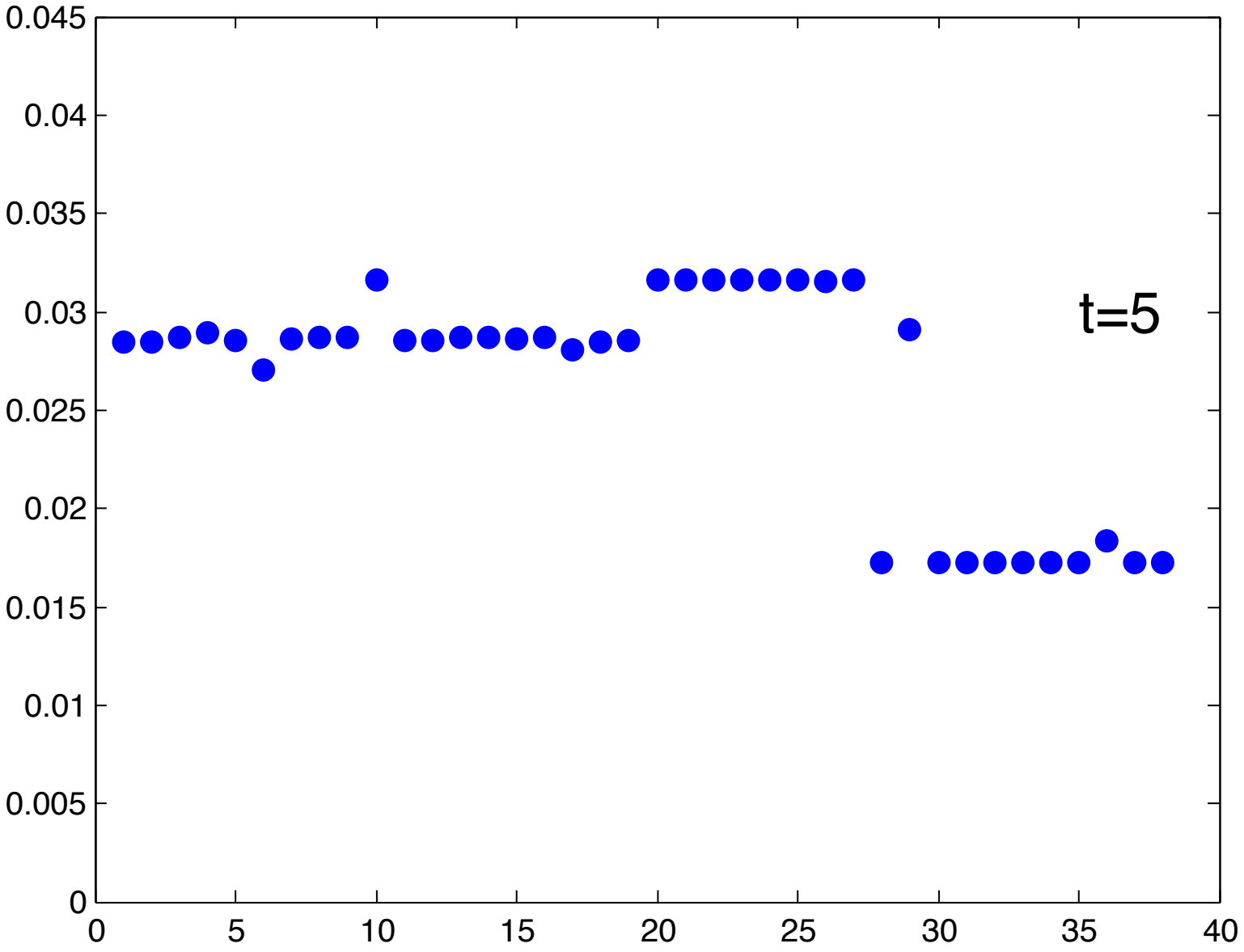
The Iterations

(Chuck Wessell, NC State)



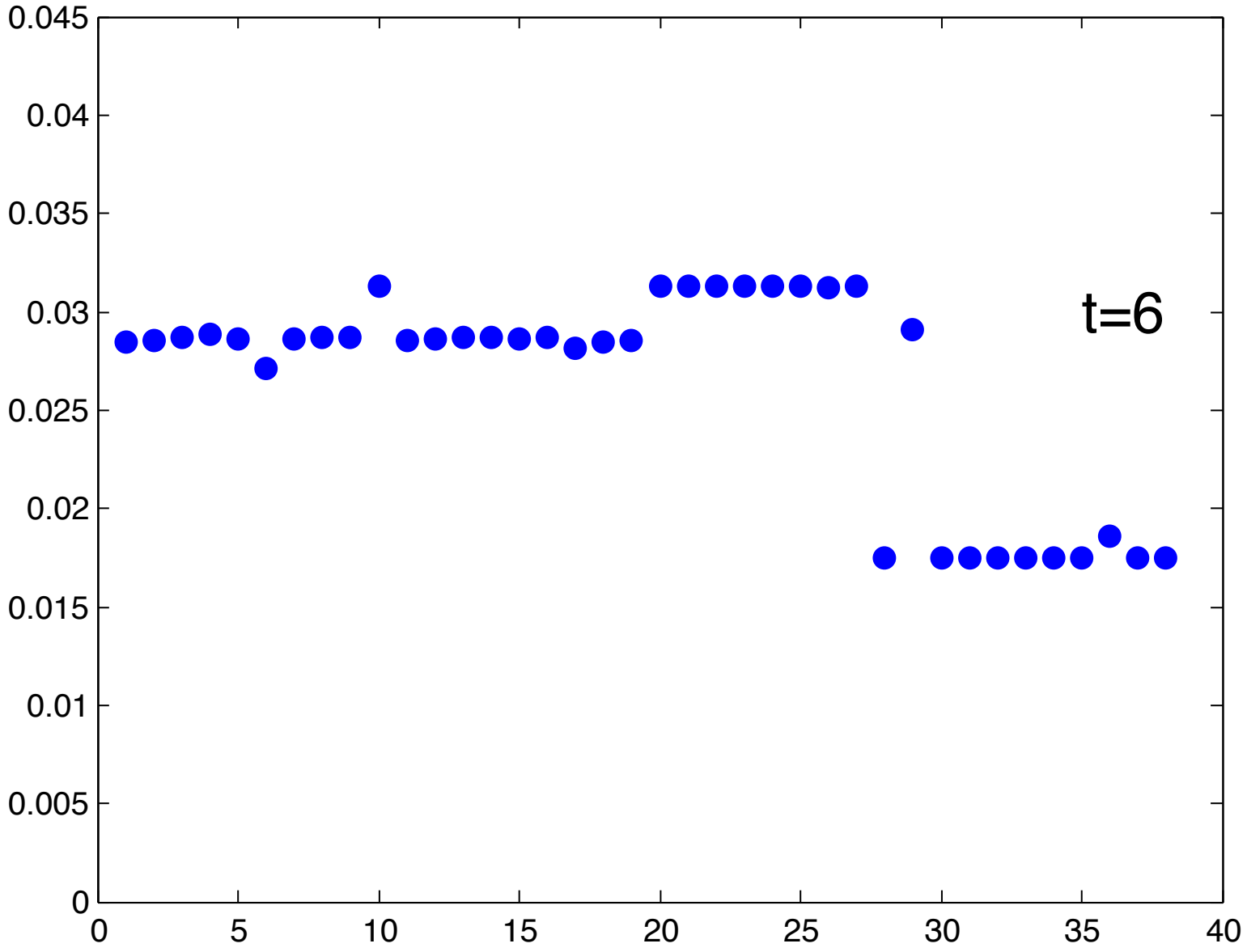
The Iterations

(Chuck Wessell, NC State)



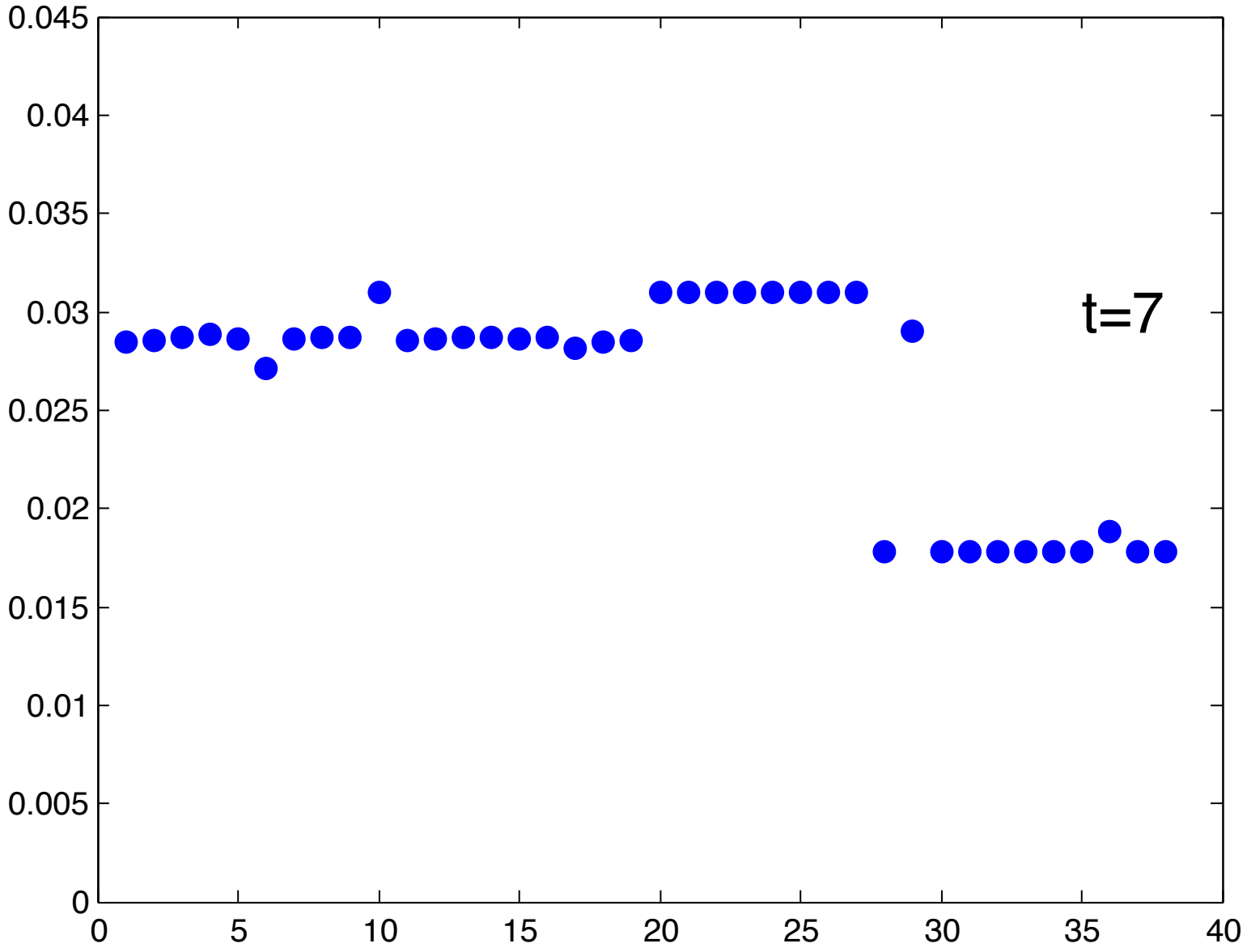
The Iterations

(Chuck Wessell, NC State)



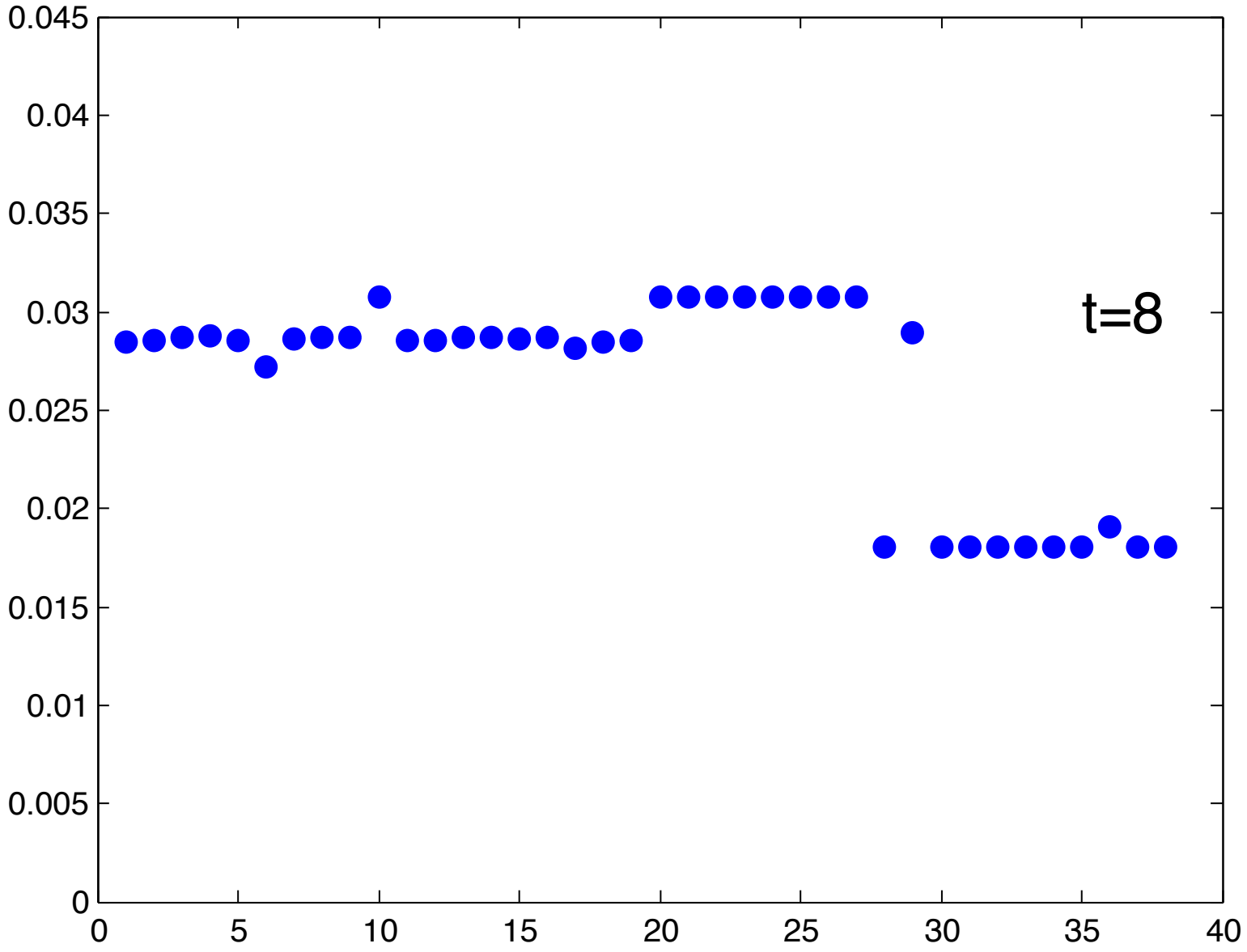
The Iterations

(Chuck Wessell, NC State)



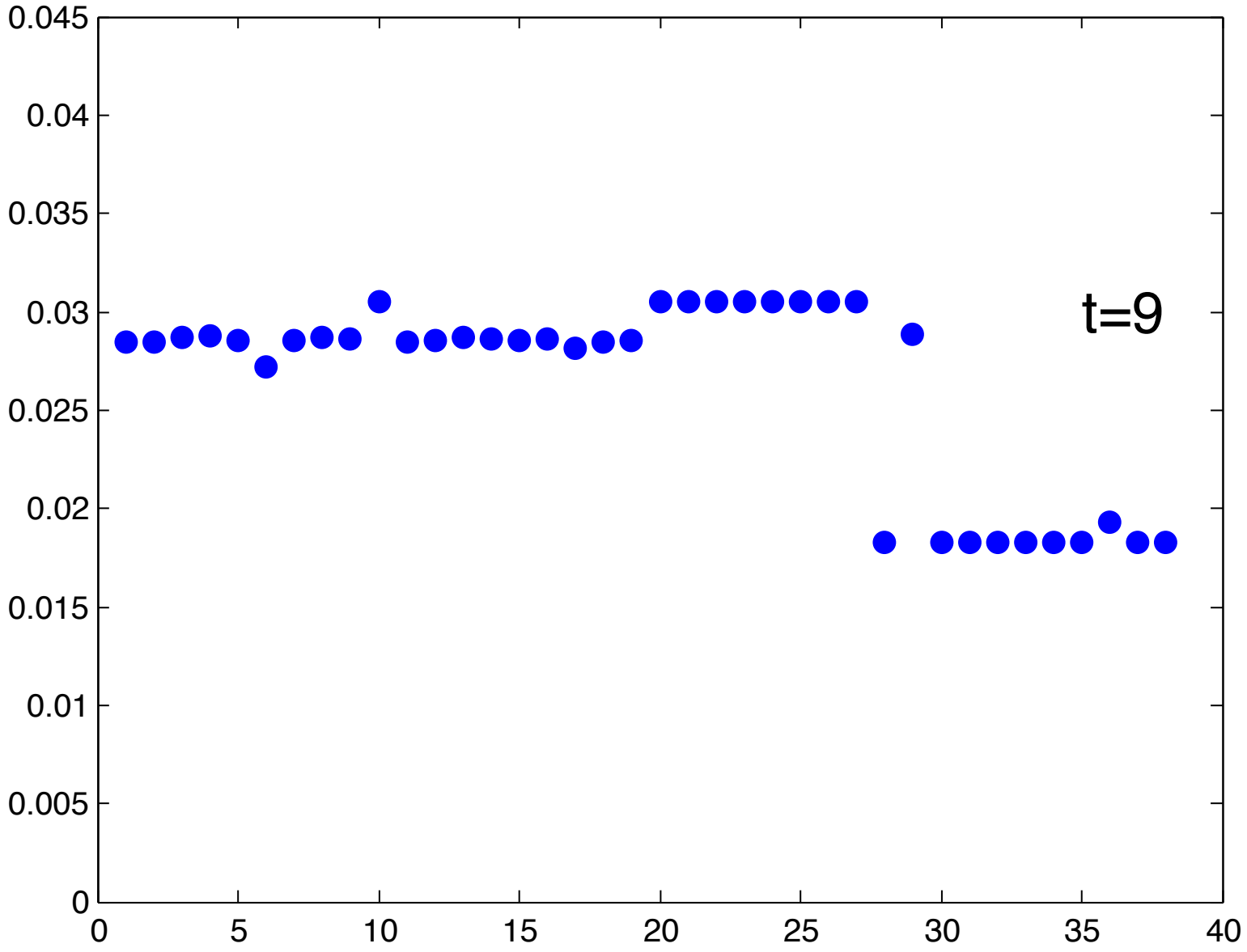
The Iterations

(Chuck Wessell, NC State)



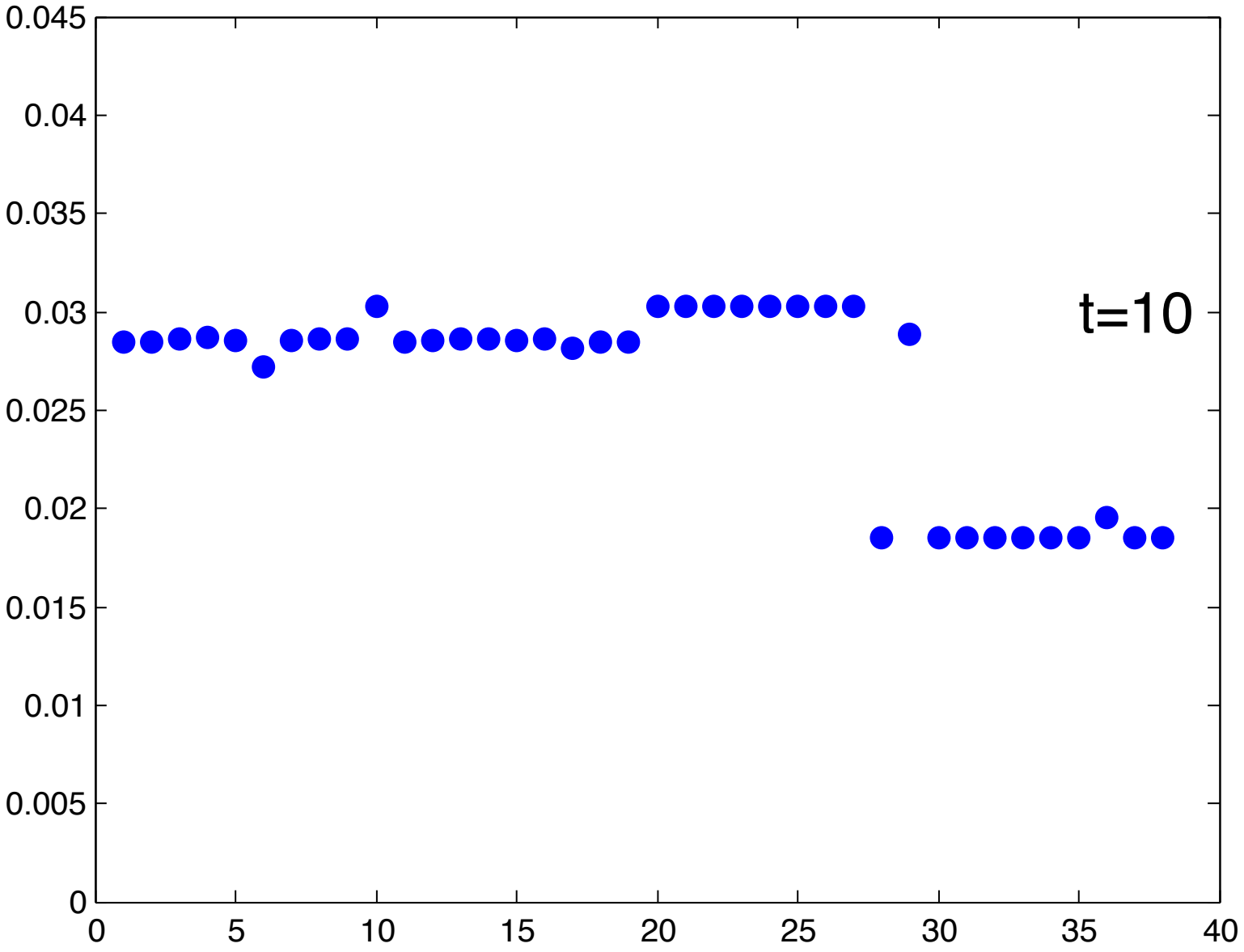
The Iterations

(Chuck Wessell, NC State)



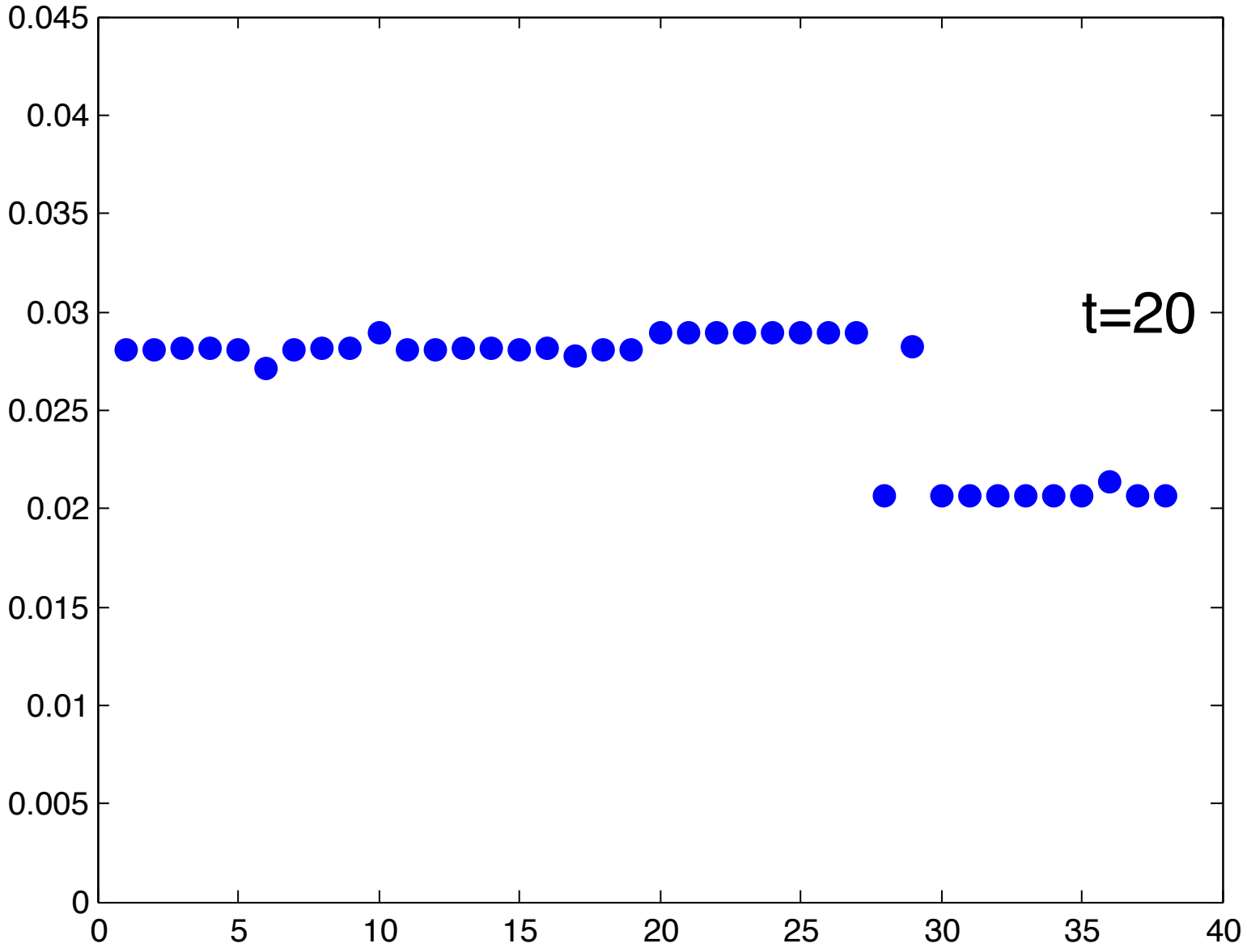
The Iterations

(Chuck Wessell, NC State)



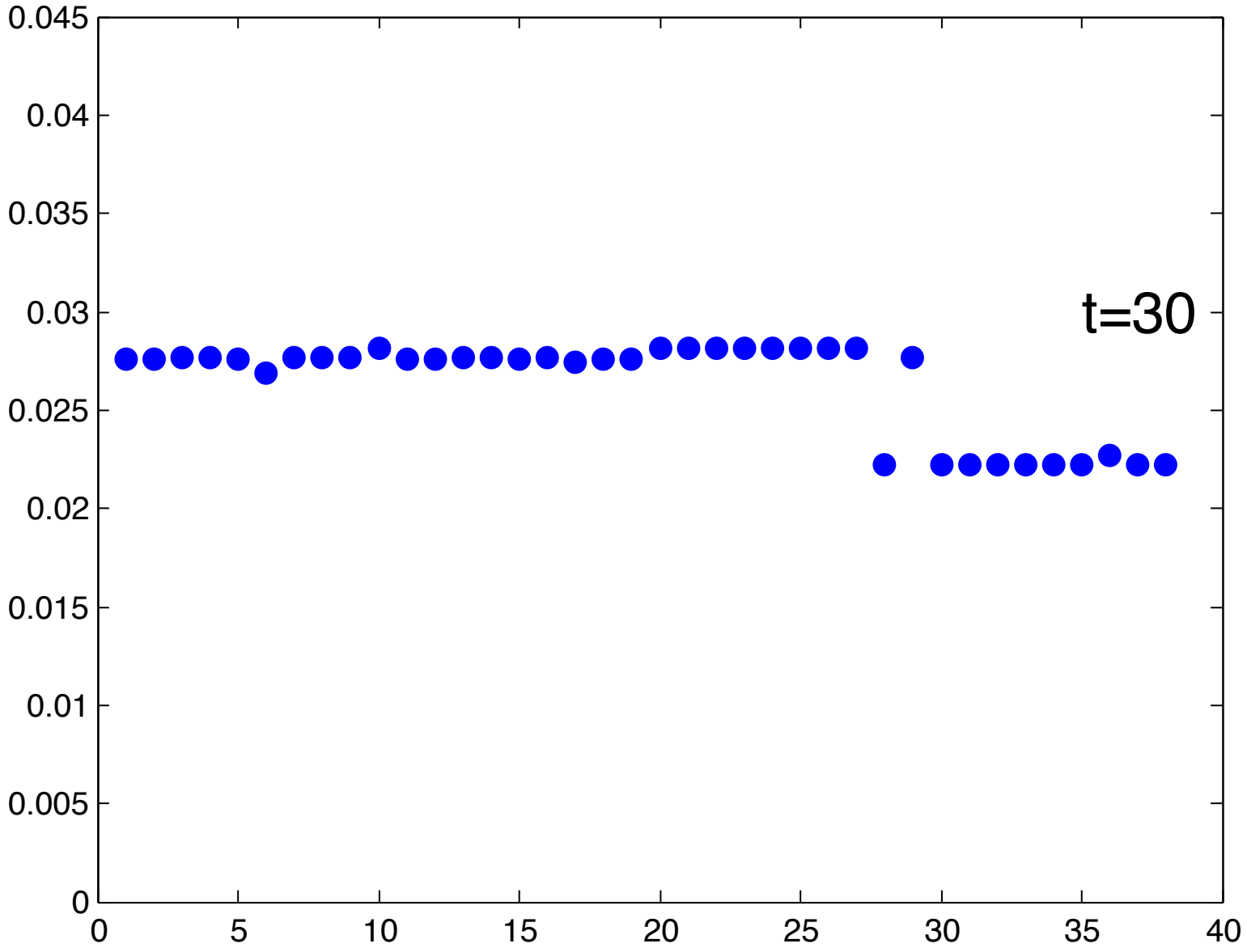
The Iterations

(Chuck Wessell, NC State)



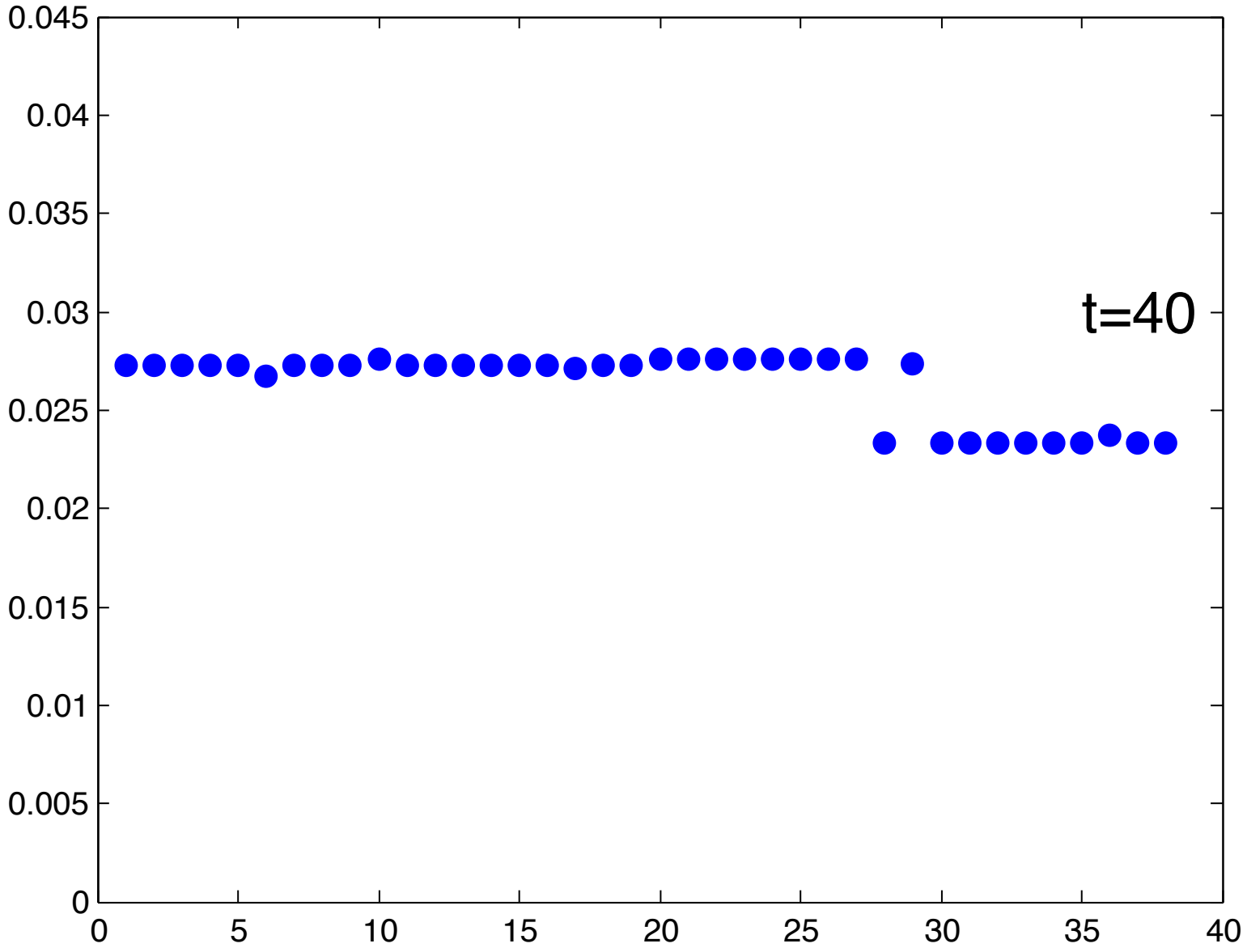
The Iterations

(Chuck Wessell, NC State)



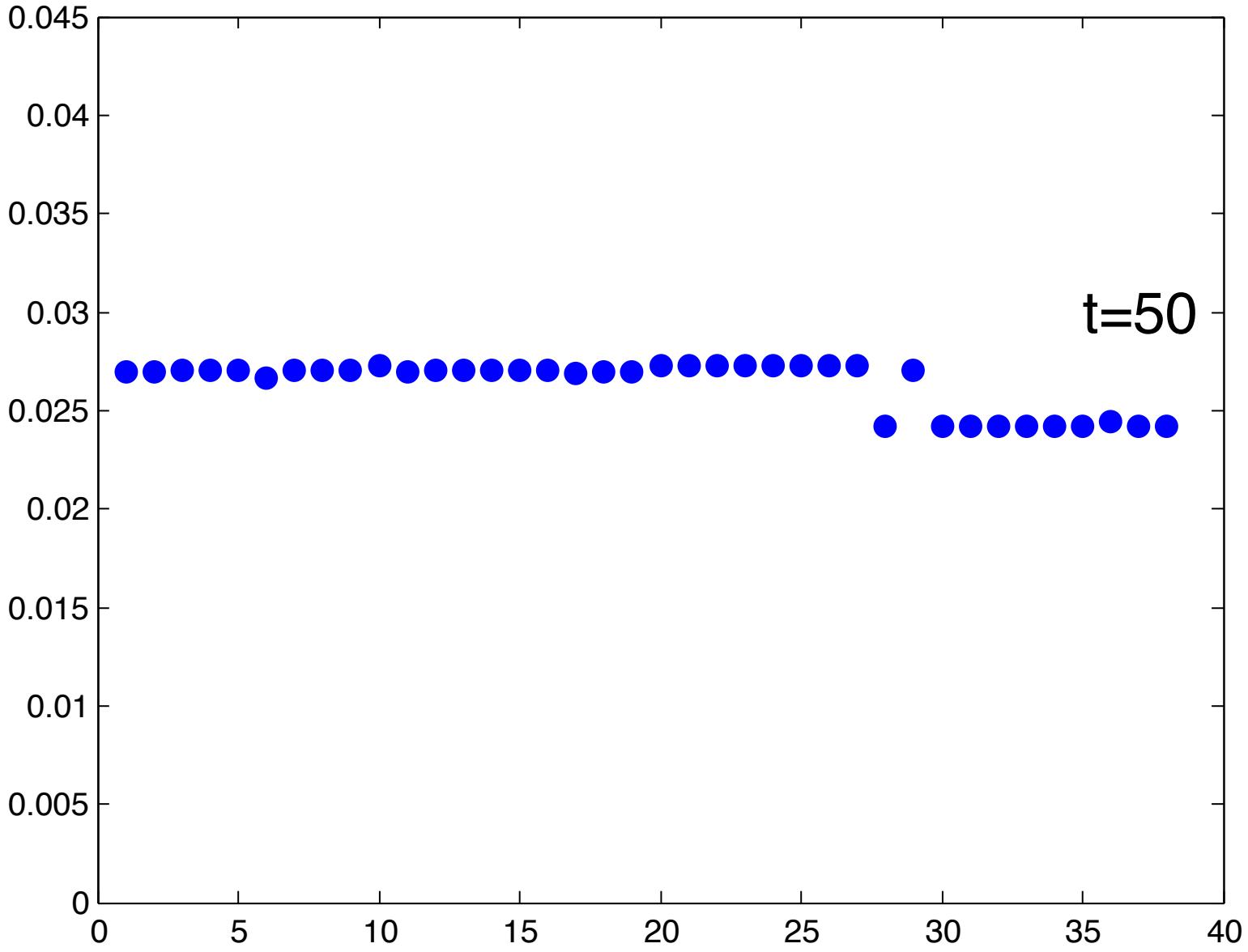
The Iterations

(Chuck Wessell, NC State)



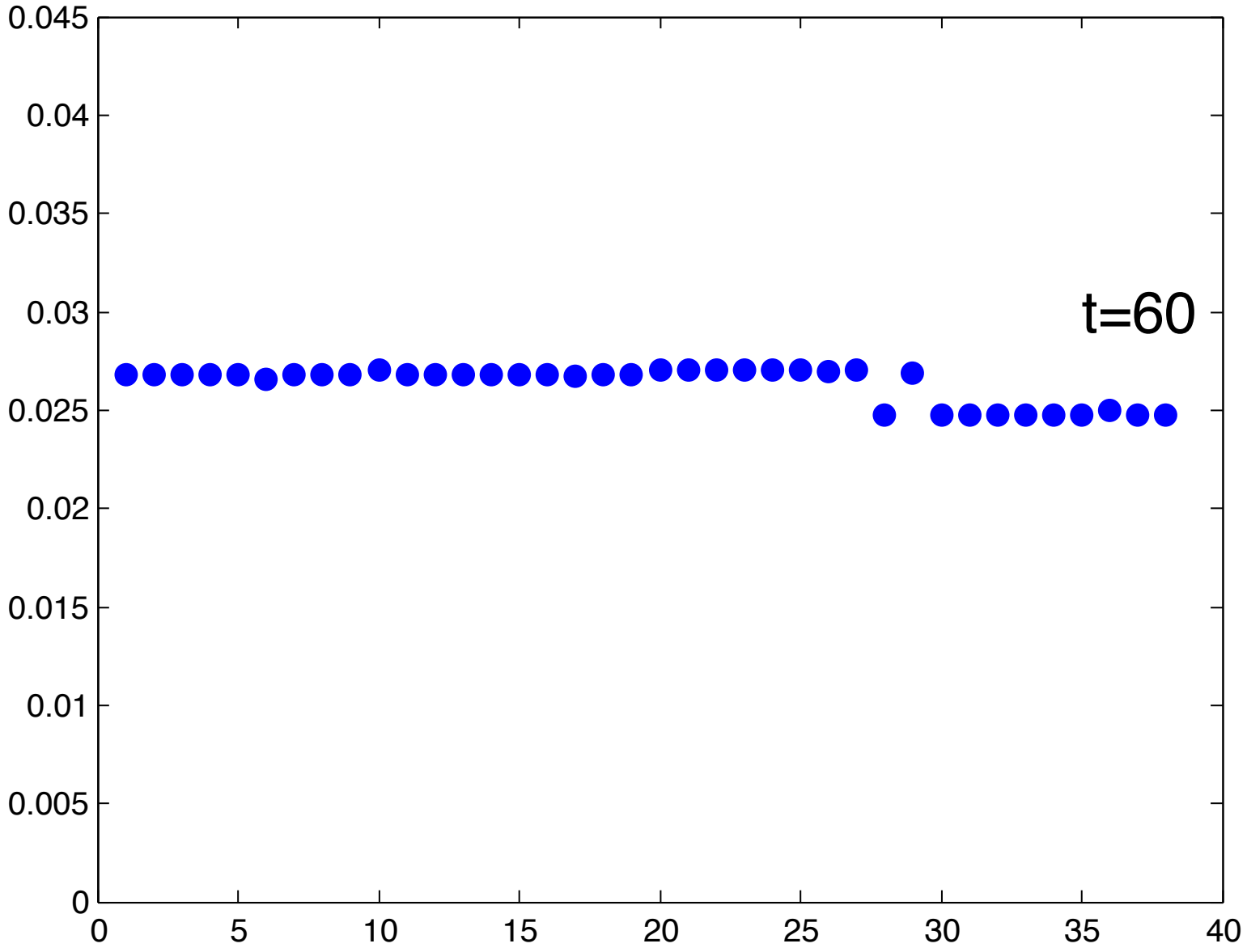
The Iterations

(Chuck Wessell, NC State)



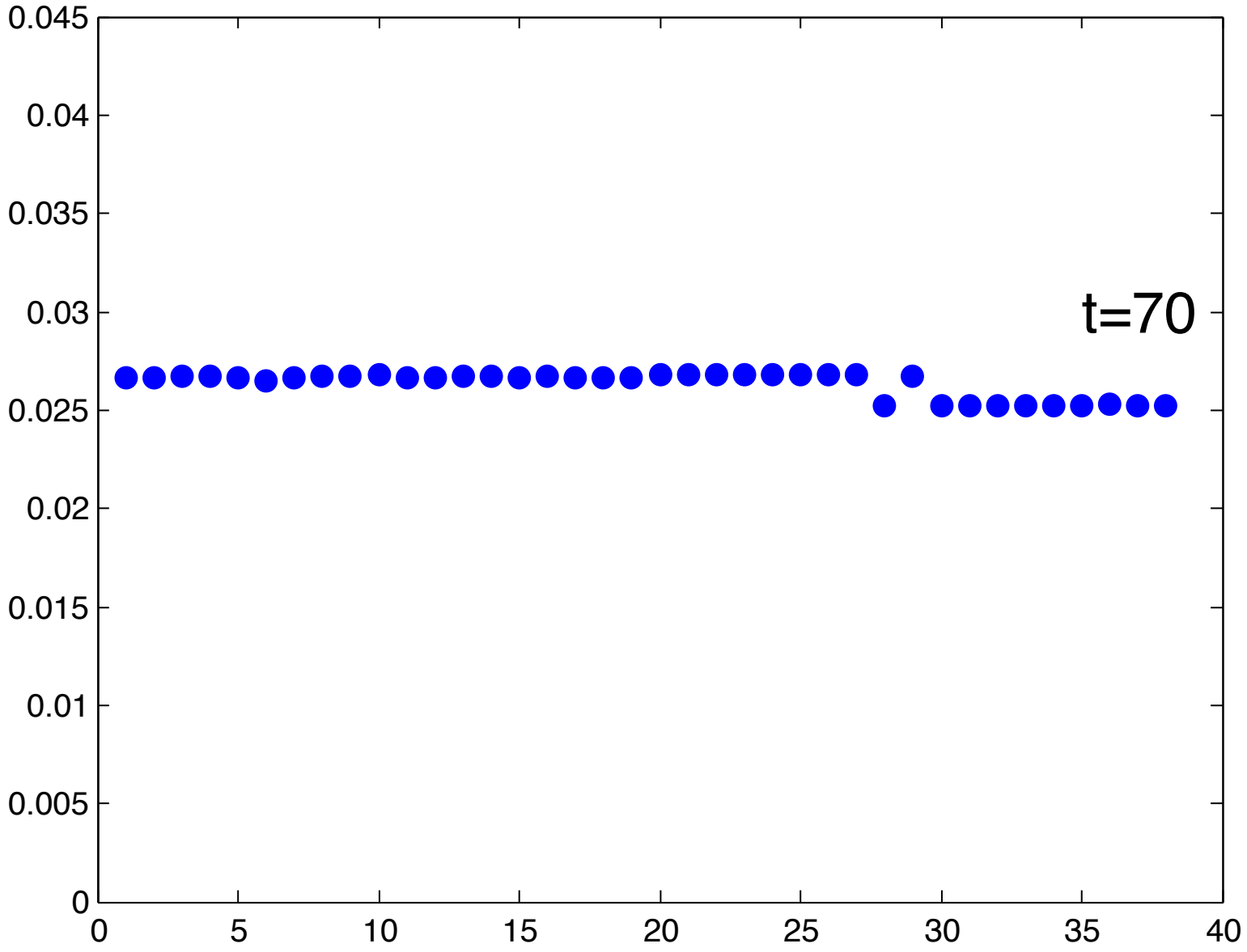
The Iterations

(Chuck Wessell, NC State)



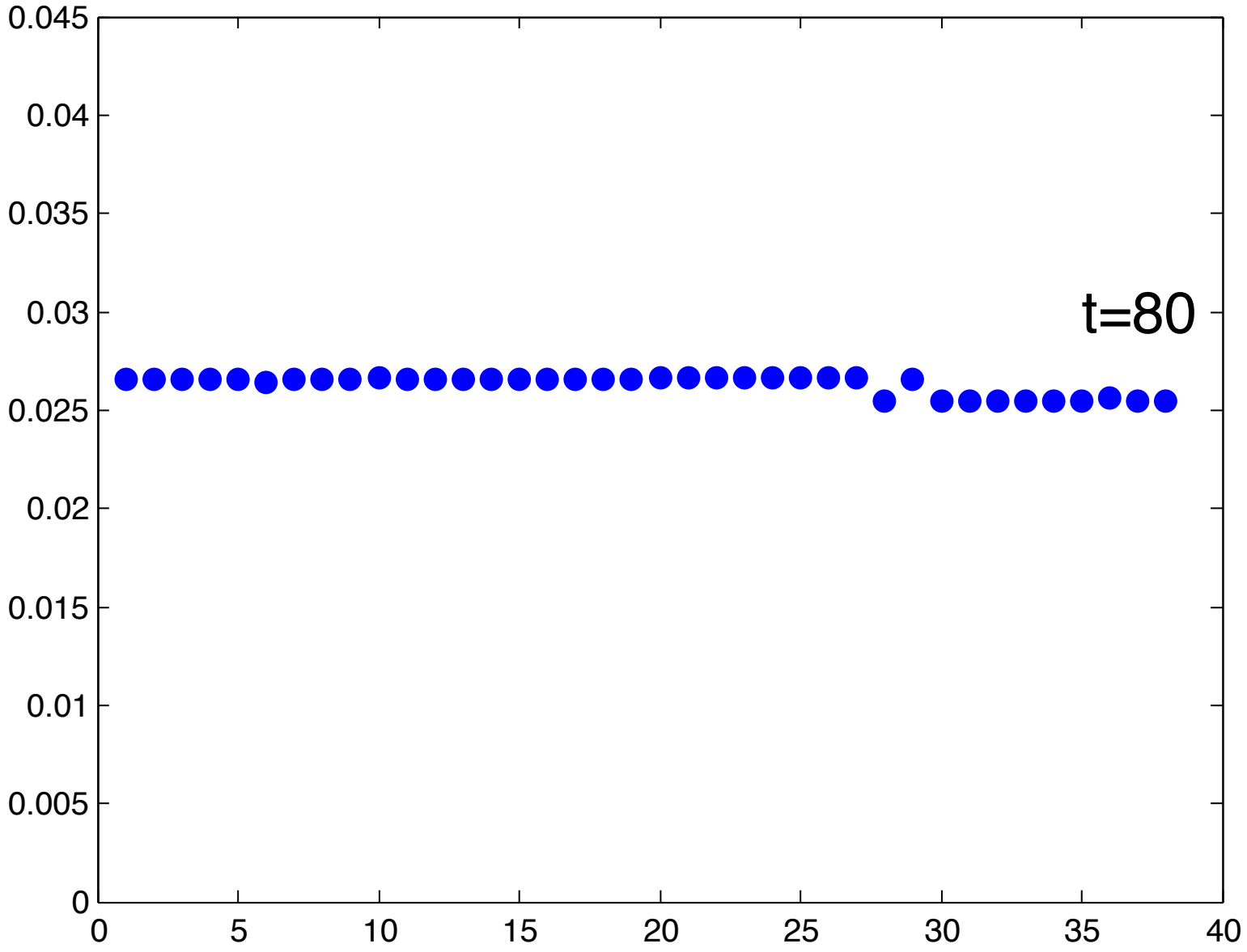
The Iterations

(Chuck Wessell, NC State)



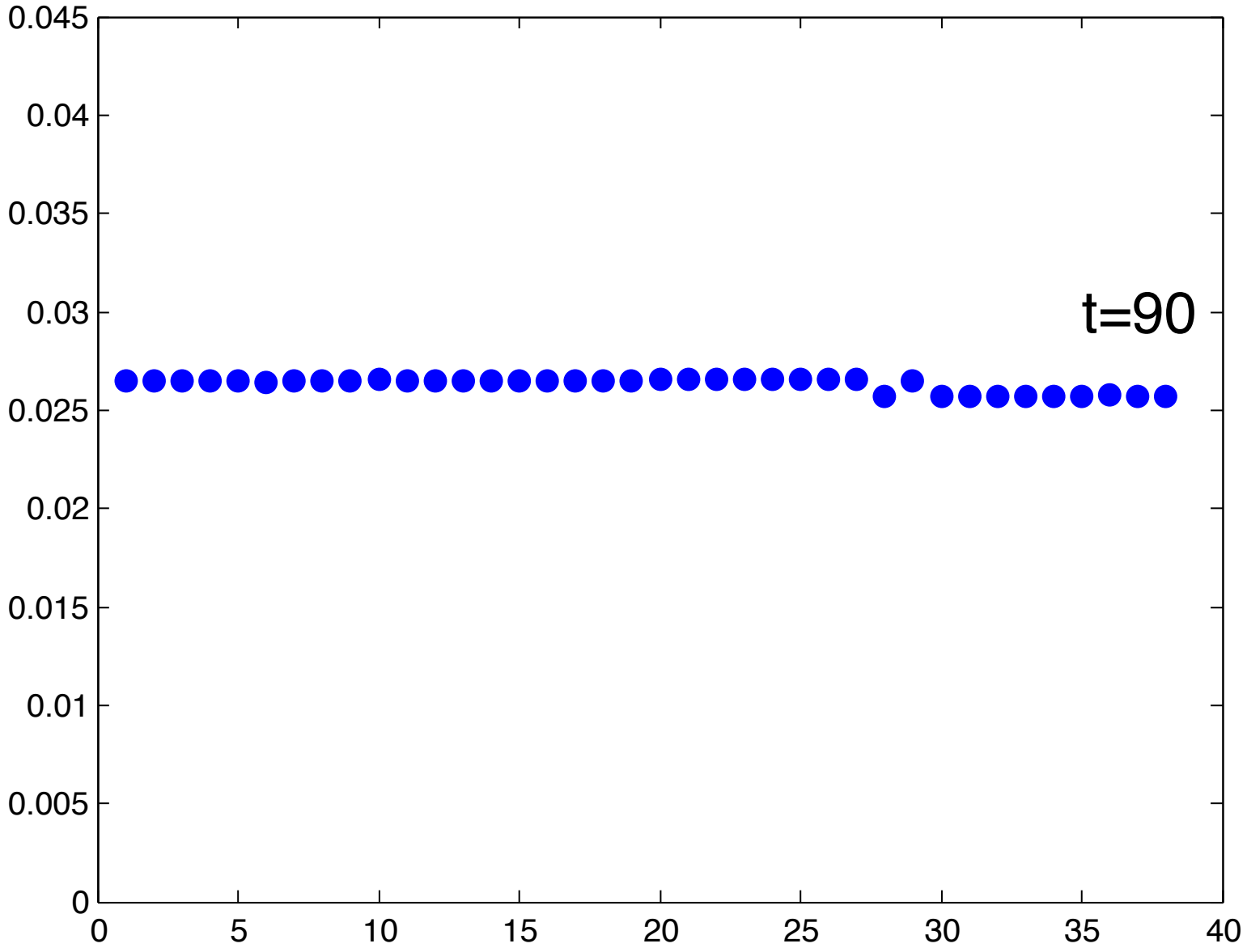
The Iterations

(Chuck Wessell, NC State)



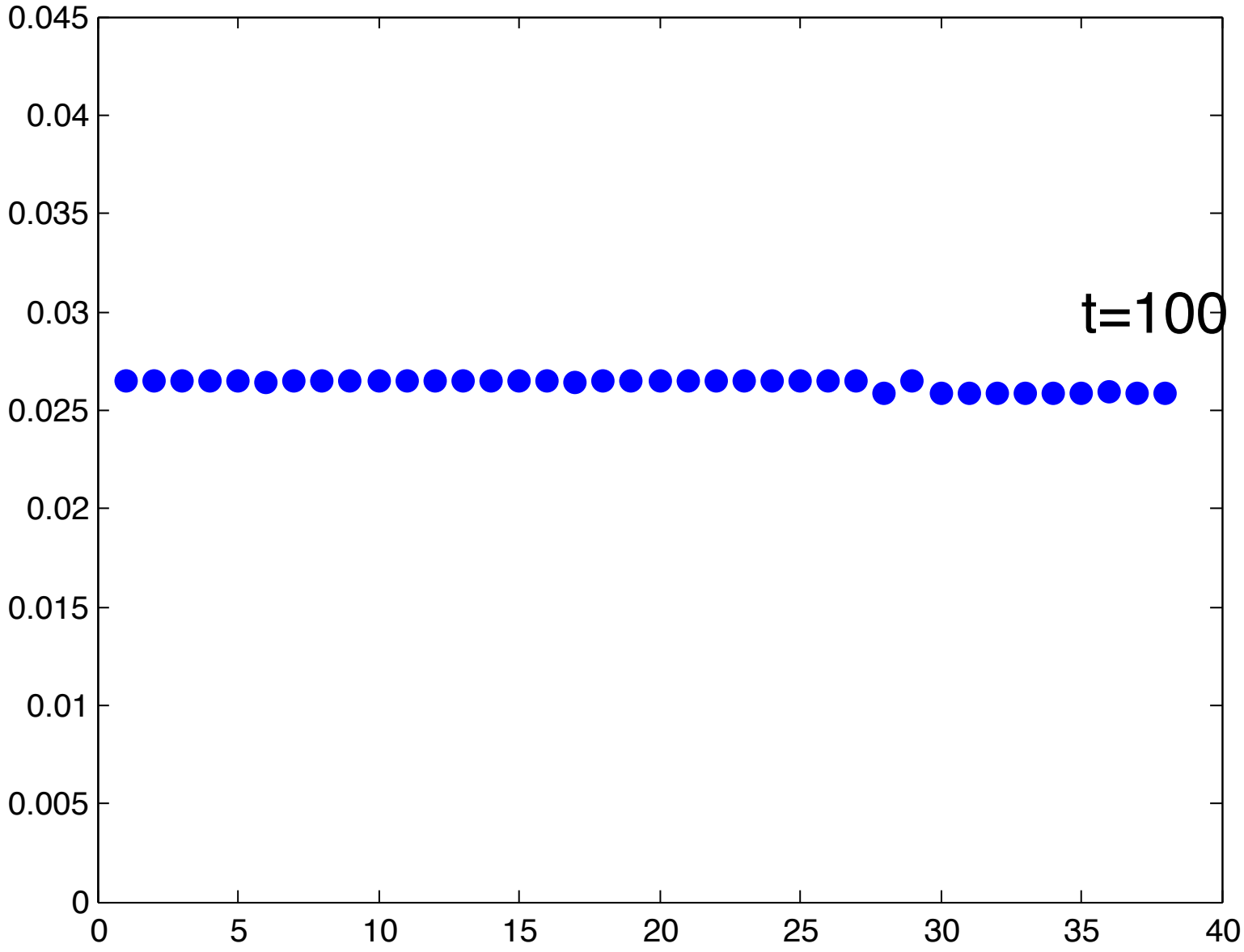
The Iterations

(Chuck Wessell, NC State)



The Iterations

(Chuck Wessell, NC State)

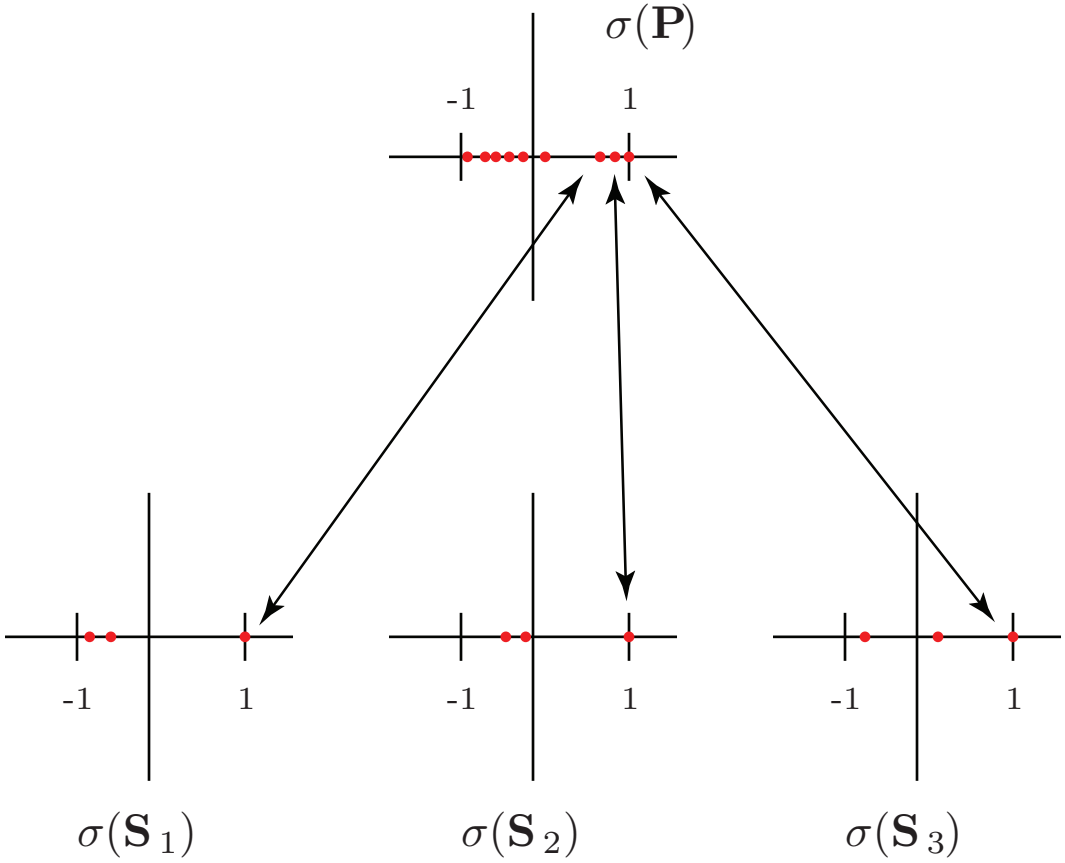


Determining The Number Of Clusters

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \mathbf{P}_{13} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \mathbf{P}_{23} \\ \mathbf{P}_{31} & \mathbf{P}_{32} & \mathbf{P}_{33} \end{bmatrix} \approx \begin{bmatrix} \mathbf{S}_1 & & \\ & \mathbf{S}_2 & \\ & & \mathbf{S}_3 \end{bmatrix}$$

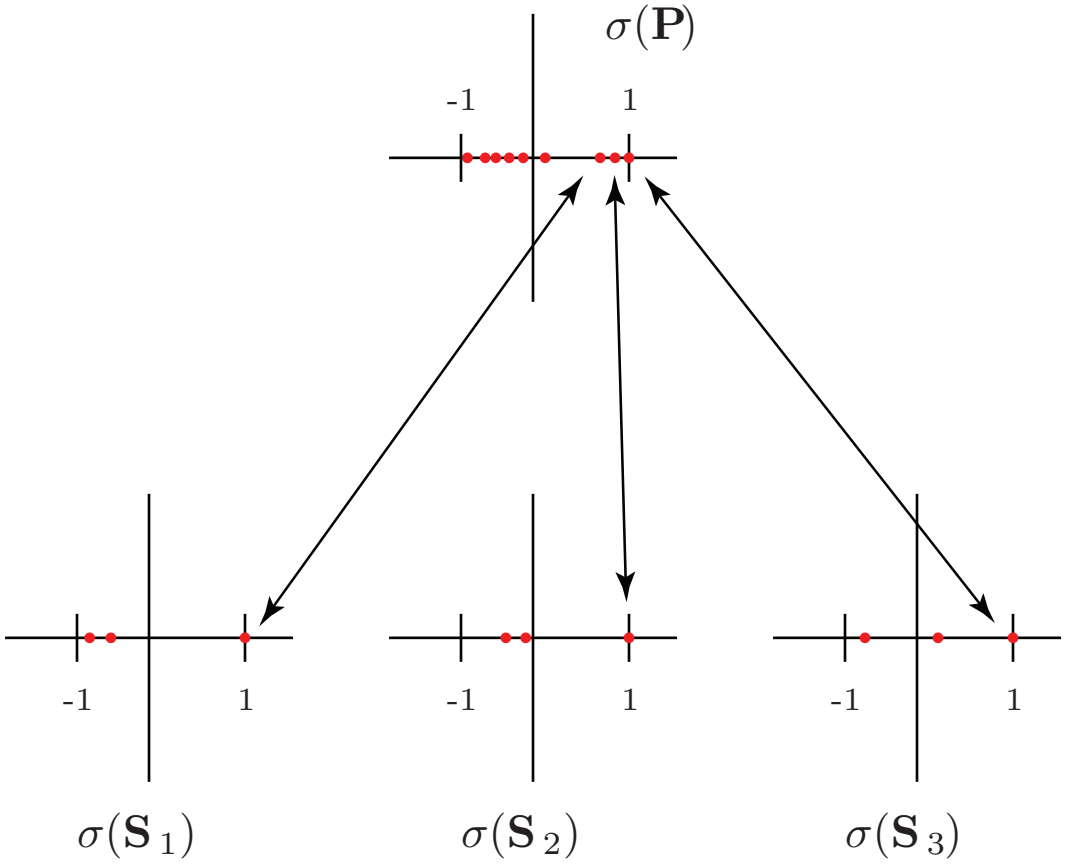
Determining The Number Of Clusters

$$\mathbf{P} = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix} \approx \begin{bmatrix} S_1 & & \\ & S_2 & \\ & & S_3 \end{bmatrix}$$



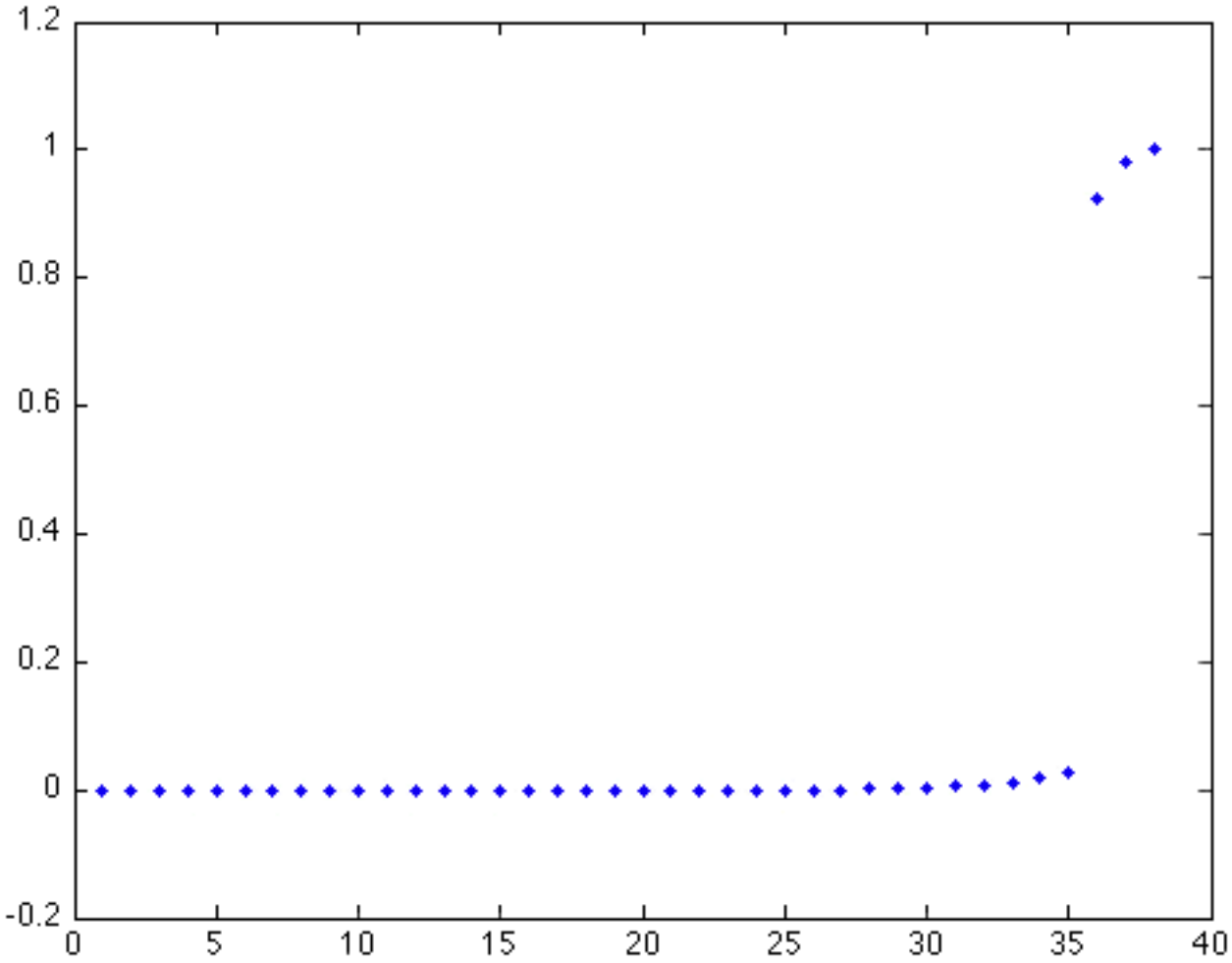
Determining The Number Of Clusters

$$\mathbf{P} = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix} \approx \begin{bmatrix} S_1 & & \\ & S_2 & \\ & & S_3 \end{bmatrix}$$



- k = # clusters = # eigenvalues near $\lambda = 1$ (determined by largest gap)

Eigenvalues For Leukemia Data



Complete Dynamic Clustering Algorithm

- Step 1. $\mathbf{A}$ (Raw data) $\longrightarrow$ $\mathbf{C} = \mathbf{C}^T$ (Similarity matrix)

  $c_{ij} = d(\mathbf{a}_i, \mathbf{a}_j)$ via similarity metric of your choice

Complete Dynamic Clustering Algorithm

Step 1. $\mathbf{A}$ (Raw data) $\longrightarrow$ $\mathbf{C} = \mathbf{C}^T$ (Similarity matrix)

$c_{ij} = d(\mathbf{a}_i, \mathbf{a}_j)$ via similarity metric of your choice

- Step 2. **Sinkhorn–Knopp Scaling**

  $\mathbf{C}$ (Similarity matrix) $\longrightarrow$ $\mathbf{P} = \mathbf{P}^T$ (Doubly stochastic)

Complete Dynamic Clustering Algorithm

Step 1. $\mathbf{A}$ (Raw data) $\longrightarrow$ $\mathbf{C} = \mathbf{C}^T$ (Similarity matrix)

$c_{ij} = d(\mathbf{a}_i, \mathbf{a}_j)$ via similarity metric of your choice

Step 2. **Sinkhorn–Knopp Scaling**

$\mathbf{C}$ (Similarity matrix) $\longrightarrow$ $\mathbf{P} = \mathbf{P}^T$ (Doubly stochastic)

- Step 3. Determine k (the number of clusters)

  k = # eigenvalues of $\mathbf{P}$ closest to 1 (Determined by largest eigengap)

Complete Dynamic Clustering Algorithm

Step 1. $\mathbf{A}$ (Raw data) $\longrightarrow$ $\mathbf{C} = \mathbf{C}^T$ (Similarity matrix)

$c_{ij} = d(\mathbf{a}_i, \mathbf{a}_j)$ via similarity metric of your choice

Step 2. **Sinkhorn–Knopp Scaling**

$\mathbf{C}$ (Similarity matrix) $\longrightarrow$ $\mathbf{P} = \mathbf{P}^T$ (Doubly stochastic)

Step 3. Determine k (the number of clusters)

$k = \#$ eigenvalues of $\mathbf{P}$ closest to 1 (Determined by largest eigengap)

- Step 4. **Initialize Markov Chain**

  Pick $\pi(0)$ to be significantly different than uniform

Complete Dynamic Clustering Algorithm

Step 1. $\mathbf{A}$ (Raw data) $\longrightarrow$ $\mathbf{C} = \mathbf{C}^T$ (Similarity matrix)

$c_{ij} = d(\mathbf{a}_i, \mathbf{a}_j)$ via similarity metric of your choice

Step 2. **Sinkhorn–Knopp Scaling**

$\mathbf{C}$ (Similarity matrix) $\longrightarrow$ $\mathbf{P} = \mathbf{P}^T$ (Doubly stochastic)

Step 3. Determine k (the number of clusters)

k = # eigenvalues of $\mathbf{P}$ closest to 1 (Determined by largest eigengap)

Step 4. **Initialize Markov Chain**

Pick $\pi(0)$ to be significantly different than uniform

- Step 5. **Observe The Chain For A Few Steps**

  $\pi(t+1) = \pi(t)\mathbf{P}$ $\quad t = 1, 2, \ldots, a$ (Until short-run stabilization)

  Order $\pi(t > a)$ and partition at the $k-1$ largest gaps

  States in each of these k segments define the k clusters

Conclusion

- Allows for dynamic visualization tools for cluster analysis

Conclusion

Allows for dynamic visualization tools for cluster analysis

- Allows the ability to select & analyze clusters in isolation

Conclusion

Allows for dynamic visualization tools for cluster analysis

Allows the ability to select & analyze clusters in isolation

- Allows tracking & analysis of selected pieces of data

Conclusion

Allows for dynamic visualization tools for cluster analysis

Allows the ability to select & analyze clusters in isolation

Allows tracking & analysis of selected pieces of data

- Works well for determining the number of clusters

Conclusion

Allows for dynamic visualization tools for cluster analysis

Allows the ability to select & analyze clusters in isolation

Allows tracking & analysis of selected pieces of data

Works well for determining the number of clusters

- Several possibilities for variations & innovations

Conclusion

Allows for dynamic visualization tools for cluster analysis

Allows the ability to select & analyze clusters in isolation

Allows tracking & analysis of selected pieces of data

Works well for determining the number of clusters

Several possibilities for variations & innovations

Remaining Questions

- Scalability

Conclusion

Allows for dynamic visualization tools for cluster analysis

Allows the ability to select & analyze clusters in isolation

Allows tracking & analysis of selected pieces of data

Works well for determining the number of clusters

Several possibilities for variations & innovations

Remaining Questions

Scalability

- Application dependency

Conclusion

Allows for dynamic visualization tools for cluster analysis

Allows the ability to select & analyze clusters in isolation

Allows tracking & analysis of selected pieces of data

Works well for determining the number of clusters

Several possibilities for variations & innovations

Remaining Questions

Scalability

Application dependency

- Sensitivity to similarity metric

Conclusion

Allows for dynamic visualization tools for cluster analysis

Allows the ability to select & analyze clusters in isolation

Allows tracking & analysis of selected pieces of data

Works well for determining the number of clusters

Several possibilities for variations & innovations

Remaining Questions

Scalability

Application dependency

Sensitivity to similarity metric

- Effectiveness of application directly to raw data

Thanks For Your Attention!