*Sebastes aleutus* (rougheyere rockfish) possibly born the same year as Darwin
Of Flour Beetles and Wikipedians
Marc Mangel
UCSC

A retro-classic
The Wikipedia

Jack Baskin School of Engineering
Biotechnology, Information Technology, Nanotechnology
The PARC Museum of Innovation
The PARC Museum of Innovation
The PARC Museum of Innovation
Tablet computer ParcPad (1992)

This notebook-sized computer with a pen-based interface was a platform for PARC research in networking, user interfaces, and work practices study. It made infrared connections but maintained constant network connectivity using a “near-field” radio system, and its electronic pen had a built-in microphone. The device also featured PARC’s Portable Common Runtime, a programming-language-independent operating system for enabling interoperability.
PhD 1994,
Alan Mackworth &
Peter Lawrence
Outline

• Wikipedia -- a bit of history

• The population biology of competition between flour beetles

• The population biology of Wikipedians

• A few conclusions about the success and failures of social collaboration networks

• Advice to the young ones
Daniel Pink’s Thought Experiment

Imagine it’s 1995.
Daniel Pink’s Thought Experiment

Imagine it’s 1995.

“I’m going to describe two new encyclopedias – one just out, the other to be launched in a few years. You have to predict which will be more successful in 2010.” “Bring it on,” she says.
Daniel Pink’s Thought Experiment

Imagine it’s 1995.

“I’m going to describe two new encyclopedias – one just out, the other to be launched in a few years. You have to predict which will be more successful in 2010.” “Bring it on,” she says.

“The first encyclopedia comes from Microsoft. As you know, Microsoft is already a large and profitable company. And with this year’s introduction of Windows 95, it’s about to become an era-defining colossus. Microsoft will fund this encyclopedia. It will pay professional writers and editors to craft articles on thousands of topics. Well-compensated managers will oversee the project to ensure it’s completed on budget and on time. Then Microsoft will sell the encyclopedia on CD-ROMs and later online.
Daniel Pink’s Thought Experiment

Imagine it’s 1995.

“I’m going to describe two new encyclopedias – one just out, the other to be launched in a few years. You have to predict which will be more successful in 2010.” “Bring it on,” she says.

“The first encyclopedia comes from Microsoft. As you know, Microsoft is already a large and profitable company. And with this year’s introduction of Windows 95, it’s about to become an era-defining colossus. Microsoft will fund this encyclopedia. It will pay professional writers and editors to craft articles on thousands of topics. Well-compensated managers will oversee the project to ensure it’s completed on budget and on time. Then Microsoft will sell the encyclopedia on CD-ROMs and later online.

“The second encyclopedia won’t come from a company. It will be created by tens of thousands of people who write and edit articles for fun. These hobbyists won’t need any special qualifications to participate. And nobody will be paid a dollar or a euro or a yen to write or edit articles. Participants will have to contribute their labor – sometimes twenty and thirty hours per week – for free. The encyclopedia itself, which will exist online, will also be free – no charge for anyone who wants to use it.”
In 1995, I doubt you could have found a single sober economist anywhere on planet Earth who would not have picked that first model as the success…
In 1995, I doubt you could have found a single sober economist anywhere on
planet Earth who would not have picked that first model as the success…

Sure, that ragtag band of volunteers might produce something. But there was no
way its product could compete with an offering from a powerful profit-driven
company. The incentives were all wrong…
In 1995, I doubt you could have a found a single sober economist anywhere on planet Earth who would not have picked that first model as the success…

Sure, that ragtag band of volunteers might produce something. But there was no way its product could compete with an offering from a powerful profit-driven company. The incentives were all wrong…

But you know how things turned out.
In 1995, I doubt you could have found a single sober economist anywhere on planet Earth who would not have picked that first model as the success…

Sure, that ragtag band of volunteers might produce something. But there was no way its product could compete with an offering from a powerful profit-driven company. The incentives were all wrong…

But you know how things turned out.

On October 31, 2009, Microsoft pulled the plug on MSN Encarta, its disc and online encyclopedia, which had been on the market for sixteen years. Meanwhile, Wikipedia – that second model – ended up becoming the largest and most popular encyclopedia in the world. Just eight years after its inception, Wikipedia had more than 13 million articles in some 260 languages, including 3 million in English alone. What happened? The conventional view of human motivation has a very hard time explaining this result.
In 1995, I doubt you could have found a single sober economist anywhere on planet Earth who would not have picked that first model as the success…

Sure, that ragtag band of volunteers might produce something. But there was no way its product could compete with an offering from a powerful profit-driven company. The incentives were all wrong…

But you know how things turned out.

On October 31, 2009, Microsoft pulled the plug on *MSN Encarta*, its disc and online encyclopedia, which had been on the market for sixteen years. Meanwhile, Wikipedia – that second model – ended up becoming the largest and most popular encyclopedia in the world. Just eight years after its inception, Wikipedia had more than 13 million articles in some 260 languages, including 3 million in English alone. What happened? The conventional view of human motivation has a very hard time explaining this result.

**Events of Jan 2012 (the Wikipedia blackout) showed what has happened**
Number of Wikipedia Editors Vs. Time

Data: B. Suh
Number of Wikipedia Editors Vs. Time

Data: B. Suh

Exponential growth
Number of Wikipedia Editors Vs. Time

Crisis! (A different talk -- “Of Seals and Wikipedians”)

Data: B. Suh
Pretty Much Every Metric Shows Exponential Growth, Saturation, and Fluctuation (Suh et al 2009)

Monthly edits as a function of time when editors are classified according to the number of lifetime edits they have made.
Connections Between Users (Nemoto and Gloor 2010)
Was the Wikipedia Guaranteed to Succeed?
Was the Wikipedia Guaranteed to Succeed?

“History is Written by the Victors” -- Winnie
Was the Wikipedia Guaranteed to Succeed?

"History is Written by the Victors" -- Winnie

If success was not guaranteed, could we predict the chance of success?
The First 18 Months Give a Very Different Picture than Guaranteed Success

![Graph showing active editors over months since start]
The First 18 Months Give a Very Different Picture than Guaranteed Success
The First 18 Months Give a Very Different Picture than Guaranteed Success

Some spectacular failed wikis
- Amazon.com
- Penguin press
- LA Times
The First 18 Months Give a Very Different Picture than Guaranteed Success

View this as a question in population biology of Active Editors

Some spectacular failed wikis
- Amazon.com
- Penguin press
- LA Times
“Population biology”, as this art-form is called, is one of the few uses of mathematics in biology to which all biologists are reconciled, and often with unaccustomed enthusiasm.

Outline

• Wikipedia -- a bit of history

• The population biology of competition between flour beetles

• The population biology of Wikipedians

• A few conclusions about the success and failures of social collaboration networks

• Advice to the young ones
A Retro-Classic

*Tribolium confusum*
(The confused flour beetle)


*Tribolium castaneum*
(The red flour beetle)

(Clemson University USDA Cooperative Extension Slide Series)
A new kind of statistics lab (IAMS)
Third Berkeley Symposium on Mathematical Statistics and Probability, Volume IV

1956

Inspired work by

VD Barnett
MS Bartlett
JC Gower
PH Leslie
The key result (for this talk): Zones of Determinate Outcome and of Indeterminate Outcome

Treatments differ by temperature and humidity.
Focal question

Given initial numbers of the two species, what will be the outcome of the competition?
In retrospect, it is perhaps informative that Park initially was not motivated by questions of stochastic events but, rather, with the demonstration that certain reasonably realistic ecological factors could affect the survival of one species over another when both were competing. It is also amusing to record, again in retrospect, that Park’s growing appreciation of indeterminacy was nurtured by the data themselves, by his association with the statisticians PH Leslie, J Neyman and MS Bartlett, who were interested, each in his own way, in the Tribolium experiments; and, finally, that Park had an increasing awareness of the fact that ecological systems are indeed stochastic -- that a population’s survival, even in the absence of environmental stochasticity, is related intimately to a component which has recently, and meaningfully, been described as ‘gambling for existence’

D Mertz, DA Cawthon and T Park. 1975. PNAS 73:1368
“Each in His Own Way” – Lotka Volterra Competition

\[
\begin{align*}
\frac{dN_1}{dt} &= r_1 N_1 \left( 1 - \frac{N_1 + \alpha_{12} N_2}{K_1} \right) \\
\frac{dN_2}{dt} &= r_2 N_2 \left( 1 - \frac{N_2 + \alpha_{21} N_1}{K_2} \right)
\end{align*}
\]

• Look at the single species isoclines

• Put them together

• Draw inferences
Zero Isoclines

\[
\frac{dN_1}{dt} = r_1 N_1 \left( 1 - \frac{N_1 + \alpha_{12} N_2}{K_1} \right)
\]
Zero Isoclines

\[
\frac{dN_1}{dt} = r_1 N_1 \left(1 - \frac{N_1 + \alpha_{12} N_2}{K_1}\right)
\]

\[
\frac{dN_1}{dt} = 0 \rightarrow N_1 + \alpha_{12} N_2 = K_1
\]
Zero Isoclines

Zero Isocline for \( N_1 \)

\[
\frac{dN_1}{dt} = r_1 N_1 \left( 1 - \frac{N_1 + \alpha_{12} N_2}{K_1} \right)
\]

\[
\frac{dN_1}{dt} = 0 \rightarrow N_1 + \alpha_{12} N_2 = K_1
\]
Zero Isoclines

\[
\frac{dN_1}{dt} = r_1 N_1 \left( 1 - \frac{N_1 + \alpha_{12} N_2}{K_1} \right)
\]

\[
\frac{dN_1}{dt} = 0 \rightarrow N_1 + \alpha_{12} N_2 = K_1
\]

Four scenarios result
Species 1 Always Wins
Species 2 Always Wins

Figure: Lou Gross, U. Tenn
Co-existence
Species 1 Sometimes Wins and Species 2 Sometimes Wins, Depending Upon Initial Conditions
Species 1 Sometimes Wins and Species 2 Sometimes Wins, Depending Upon Initial Conditions

This is beginning to look a lot like flour beetles
Species 1 Sometimes Wins and Species 2 Sometimes Wins, Depending Upon Initial Conditions

This is beginning to look a lot like flour beetles

Fluctuations at an unstable steady state
The Differential Equations as the Conditional Means Of a Stochastic Process (Neyman, Park and Scott; Leslie and Gower; Bartlett, Leslie and Gower; Barnett)

\[ X_1(t) = \text{Adults of } T. \text{castaneum} \]

\[ X_2(t) = \text{Adults of } T. \text{confusum} \]

Set

\[ \Delta X_i = X_i(t + \Delta t) - X_i(t) \]

Then, given \( X_i(t) = x_i \)

\[ \Pr\{\Delta X_1 = 1, \Delta X_2 = 0 \mid \text{change} \} = \frac{x_1 \lambda_1}{k} \]

\[ \Pr\{\Delta X_1 = -1, \Delta X_2 = 0 \mid \text{change} \} = \frac{x_1 \mu_1}{k} \quad k = x_1 (\lambda_1 + \mu_1) + x_2 (\lambda_2 + \mu_2) \]

\[ \Pr\{\Delta X_1 = 0, \Delta X_2 = 1 \mid \text{change} \} = \frac{x_2 \lambda_2}{k} \]

\[ \Pr\{\Delta X_1 = 0, \Delta X_2 = -1 \mid \text{change} \} = \frac{x_2 \mu_2}{k} \]
\begin{align*}
\Pr\{\Delta X_1 = 1, \Delta X_2 = 0 \mid \text{change}\} &= \frac{x_1 \lambda_1}{k} \\
\Pr\{\Delta X_1 = -1, \Delta X_2 = 0 \mid \text{change}\} &= \frac{x_1 \mu_1}{k} \\
\Pr\{\Delta X_1 = 0, \Delta X_2 = 1 \mid \text{change}\} &= \frac{x_2 \lambda_2}{k} \\
\Pr\{\Delta X_1 = 0, \Delta X_2 = -1 \mid \text{change}\} &= \frac{x_2 \mu_2}{k}
\end{align*}

\begin{align*}
k &= x_1 (\lambda_1 + \mu_1) + x_2 (\lambda_2 + \mu_2) \\
\lambda_1 &= 0.11 - 0.0007 x_1 - 0.01 x_2 \\
\lambda_2 &= 0.08 - 0.0007 x_1 - 0.007 x_2 \\
\mu_1 &= 0.01 \\
\mu_2 &= 0.005
\end{align*}

Suitable for rapid Monte Carlo simulation even 50 years ago

Deterministic dynamics as an emergent property (Mangel, TPB 1994)
Deterministic dynamics with a saddle point perturbed by noise

\[
\frac{dx}{dt} = x(1 + a - x - ay) \\
\frac{dy}{dt} = y(1 + a - y - ax)
\]

(1,1) is a steady state; stable if \( a < 1 \) [co-existence] and unstable if \( a > 1 \) [extinction of one species]
Make the dynamics stochastic

\[ dX = X(1 + a - X - aY) + \sqrt{\varepsilon}dB_1 \]

\[ dY = Y(1 + a - Y - aX) + \sqrt{\varepsilon}dB_2 \]
Make the dynamics stochastic

\[dX = X(1 + a - X - aY) + \sqrt{\varepsilon} dB_1\]
\[dY = Y(1 + a - Y - aX) + \sqrt{\varepsilon} dB_2\]

• Surround the separatrix with a band
Make the dynamics stochastic

\[ dX = X(1 + a - X - aY) + \sqrt{\varepsilon}dB_1 \]
\[ dY = Y(1 + a - Y - aX) + \sqrt{\varepsilon}dB_2 \]

- Surround the separatrix with a band

- Define \( u(x_1, x_2) = \text{Pr}\{\text{exit the band towards the y-axis} | X_i(0) = x_i\} \)
Make the dynamics stochastic

\[ dX = X(1 + a - X - aY) + \sqrt{\varepsilon}dB_1 \]
\[ dY = Y(1 + a - Y - aX) + \sqrt{\varepsilon}dB_2 \]

• Surround the separatrix with a band

• Define
  \[ u(x_1, x_2) = \Pr\{\text{exit the band towards the y-axis}|X_i(0) = x_i\} \]

• Derive an equation for \( u(x) \), construct its solution, and interpret
When Done for the Flour Beetle Birth and Death Dynamics, Use A Central Limit Theorem for Population Processes

\[ dX_1 = X_1(\lambda_1(X_1, X_2) - \mu_1(X_1, X_2)) + \sum_j \epsilon a_{1j}(X_1, X_2) dB_j \]

\[ dX_2 = X_2(\lambda_2(X_1, X_2) - \mu_2(X_1, X_2)) + \sum_j \epsilon a_{2j}(X_1, X_2) dB_j \]

Here \( \epsilon \sim 1/\text{Steady state population size} \). With the previous stochastic assumptions

\[ \epsilon a(x_1, x_2) = \begin{pmatrix}
\frac{\epsilon x_1(\lambda_1(x_1, x_2) + \mu_1(x_1, x_2))}{k} & 0 \\
0 & \frac{\epsilon x_2(\lambda_2(x_1, x_2) + \mu_2(x_1, x_2))}{k}
\end{pmatrix} \]
Results of the Plan: Contours of Indeterminancy and Analogies with Diffraction

\[ u(x_1,x_2) = \Pr\{\text{exit the band towards the y-axis} | X_i(0) = x_i \} \]
Another Application: Peak Shifts Produced by Correlated Response to Selection


http://blogs.discovermagazine.com/gnxp/tag/fitness-landscape/
Outline

• Wikipedia -- a bit of history

• The population biology of competition between flour beetles

• The population biology of Wikipedians

• A few conclusions about the success and failures of social collaboration networks

• Advice to the young ones
The Full Trajectory of Editors Requires a Nonlinear Stochastic Differential Equation

\[ dW = r(W) \, dt + \sigma(W) \, dB \]

with \( r(W) \) and \( \sigma(W) \) to be determined,
The Full Trajectory of Editors Requires a Nonlinear Stochastic Differential Equation

\[ dW = r(W)\,dt + \sigma(W)\,dB_W \]

with \( r(W) \) and \( \sigma(W) \) to be determined,

But maybe a simpler model will do at the start….
For a Wiki to Succeed, the Origin (or some point above it) Must Be An Unstable Steady State
For a Wiki to Succeed, the Origin (or some point above it) Must Be An Unstable Steady State

Given the number of editors at 18 months, say, what is the probability of crossing some large number of editors before falling back to 0?
For a Wiki to Succeed, the Origin (or some point above it) Must Be An Unstable Steady State

Given the number of editors at 18 months, say, what is the probability of crossing some large number of editors before falling back to 0?

This is the flour beetle question!
For a Wiki to Succeed, the Origin (or some point above it) Must Be An Unstable Steady State

Given the number of editors at 18 months, say, what is the probability of crossing some large number of editors before falling back to 0?

This is the flour beetle question!

\[ dW = r(W)dt + \sigma(W) dB_W \]
For a Wiki to Succeed, the Origin (or some point above it) Must Be An Unstable Steady State

Given the number of editors at 18 months, say, what is the probability of crossing some large number of editors before falling back to 0?

This is the flour beetle question!

\[
dW = r(W)dt + \sigma(W)dB_W
\]

“In the right hands, integration by parts and Taylor expansion are still among the most powerful tools of applied mathematics” - Mark Kac
Our Approach. Step 1, use a Simple Model

\[ dW = (r_0 + r_1W)dt + (\sigma_0 + \sigma_1W)dB_W \]
Our Approach. Step 1, use a Simple Model

\[ dW = (r_0 + r_1 W)dt + (\sigma_0 + \sigma_1 W)dB_W \]

Step 2. Get a Crack at the Parameters with AIC

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC Weight ()</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) ( \mu = r_0, \sigma = \sigma_0 )</td>
<td>0.36</td>
</tr>
<tr>
<td>2) ( \mu = r_0, \sigma = \sigma_0 + \sigma_1 W )</td>
<td>0.32</td>
</tr>
<tr>
<td>3) ( \mu = r_1 W, \sigma = \sigma_0 )</td>
<td>0.12</td>
</tr>
<tr>
<td>4) ( \mu = r_0 + r_1 W, \sigma = \sigma_0 )</td>
<td>0.098</td>
</tr>
<tr>
<td>5) ( \mu = r_0 + r_1 W, \sigma = \sigma_0 + \sigma_1 W )</td>
<td>0.061</td>
</tr>
</tbody>
</table>

97% of the AIC weight

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC Weight ()</th>
</tr>
</thead>
<tbody>
<tr>
<td>6) ( \mu = r_1 W, \sigma = \sigma_0 + \sigma_1 W )</td>
<td>0.012</td>
</tr>
<tr>
<td>7) ( \mu = r_0, \sigma = \sigma_1 W )</td>
<td>0.011</td>
</tr>
<tr>
<td>8) ( \mu = r_1 W, \sigma = \sigma_1 W )</td>
<td>0.004</td>
</tr>
<tr>
<td>9) ( \mu = r_0 + r_1 W, \sigma = \sigma_1 W )</td>
<td>0.0029</td>
</tr>
</tbody>
</table>
Step 3. Define

\[ u(w) = \text{Prob}\{\text{hit a large number of editors, } w_u, \text{ before 0 given that there are currently } w \text{ editors}\} \]
Step 3. Define

\[ u(w) = \text{Prob}\{\text{hit a large number of editors, } w_u, \text{ before } 0 \text{ given that there are currently } w \text{ editors}\} \]

Apply the law of total probability

\[ u(w) = E_{dW} \{u(w + dW)\} \]
Step 3. Define

\[ u(w) = \text{Prob}\{\text{hit a large number of editors, } w_u, \text{ before 0 given that there are currently } w \text{ editors} \} \]

Apply the law of total probability

\[ u(w) = E_{dW} \{ u(w + dW) \} \]

Taylor expand

\[ u(w) = E_{dW} \{ u(w) + \frac{\partial u}{\partial w} dW + \frac{1}{2} \frac{\partial^2 u}{\partial w^2} dW^2 + o(dW^2) \} \]
Step 3. Define

\[ u(w) = \text{Prob}\{\text{hit a large number of editors, } w_u, \text{ before 0 given that there are currently } w \text{ editors}\} \]

Apply the law of total probability

\[ u(w) = E_{dW} \{u(w + dW)\} \]

Taylor expand

\[ u(w) = E_{dW} \{u(w) + \frac{\partial u}{\partial w} dW + \frac{1}{2} \frac{\partial^2 u}{\partial w^2} dW^2 + o(dW^2)\} \]

Note that

\[ E\{dW \mid W(t) = w\} = (r_0 + r_1 w) dt + o(dt) \]
\[ E\{dW^2 \mid W(t) = w\} = (\sigma_0 + \sigma_1 w)^2 dt + o(dt) \]
Step 3 Continued…So

\[ u(w) = E_{dw} \{ u(w) + \frac{\partial u}{\partial w} dW + \frac{1}{2} \frac{\partial^2 u}{\partial w^2} dW^2 + o(dW^2) \} \]
Step 3 Continued…So

\[ u(w) = E_{dw} \{ u(w) + \frac{\partial u}{\partial w} dW + \frac{1}{2} \frac{\partial^2 u}{\partial w^2} dW^2 + o(dW^2) \} \]

Implies that

\[ u(w) = u(w) + (r_0 + r_1 w) \frac{\partial u}{\partial w} dt + \frac{(\sigma_0 + \sigma_1 w)^2}{2} \frac{\partial^2 u}{\partial w^2} dt + o(dt) \]
Step 3 Continued…So

\[ u(w) = E_{dW} \{ u(w) + \frac{\partial u}{\partial w} dW + \frac{1}{2} \frac{\partial^2 u}{\partial w^2} dW^2 + o(dW^2) \} \]

Implies that

\[ u(w) = u(w) + (r_0 + r_1 w) \frac{\partial u}{\partial w} dt + \frac{(\sigma_0 + \sigma_1 w)^2}{2} \frac{\partial^2 u}{\partial w^2} dt + o(dt) \]

Subtract \( u(w) \) from both sides, divide by \( dt \), and let \( dt \rightarrow 0 \)

\[ 0 = (r_0 + r_1 w) \frac{\partial u}{\partial w} + \frac{(\sigma_0 + \sigma_1 w)^2}{2} \frac{\partial^2 u}{\partial w^2} \]
Step 3 Continued...So

\[ u(w) = E_{dw} \{ u(w) + \frac{\partial u}{\partial w} dW + \frac{1}{2} \frac{\partial^2 u}{\partial w^2} dW^2 + o(dW^2) \} \]

Implies that

\[ u(w) = u(w) + (r_0 + r_1 w) \frac{\partial u}{\partial w} dt + \frac{(\sigma_0 + \sigma_1 w)^2}{2} \frac{\partial^2 u}{\partial w^2} dt + o(dt) \]

Subtract \( u(w) \) from both sides, divide by \( dt \), and let \( dt \rightarrow 0 \)

\[ 0 = (r_0 + r_1 w) \frac{\partial u}{\partial w} + \frac{(\sigma_0 + \sigma_1 w)^2}{2} \frac{\partial^2 u}{\partial w^2} \]

With boundary conditions

\[ u(0) = 0; u(w_u) = 1 \]
Step 3 Continued…Solve the ODE in general

\[ u(w) = \frac{\int_{0}^{w} \exp \left[ -2 \int_{0}^{y} \frac{r_{0} + r_{1}x}{(\sigma_{0} + \sigma_{1}x)^{2}} \, dx \right] \, dy}{\int_{0}^{w} \exp \left[ -2 \int_{0}^{y} \frac{r_{0} + r_{1}x}{(\sigma_{0} + \sigma_{1}x)^{2}} \, dx \right] \, dy} \]
Step 3 Continued... Solve the ODE in general

\[
\begin{align*}
\mathcal{u}(w) &= \frac{\int_{0}^{w} \exp \left[ -2 \int_{0}^{y} \frac{r_0 + r_1 x}{(\sigma_0 + \sigma_1 x)^2} \, dx \right] \, dy}{\int_{0}^{w} \exp \left[ -2 \int_{0}^{y} \frac{r_0 + r_1 x}{(\sigma_0 + \sigma_1 x)^2} \, dx \right] \, dy} \\
\mathcal{u}(w) &= \frac{1 - e^{-\frac{-2r_0}{\sigma_0^2}w}}{1 - e^{-\frac{-2r_0}{\sigma_0^2}w}}
\end{align*}
\]

Model 1
Step 3 Continued…Solve the ODE in general

\[
\begin{align*}
  u(w) &= \frac{\int_{0}^{w} \exp \left[ -2 \int_{0}^{y} \frac{r_0 + r_1 x}{(\sigma_0 + \sigma_1 y)^2} \, dx \right] \, dy}{\int_{0}^{w} \exp \left[ -2 \int_{0}^{y} \frac{r_0 + r_1 x}{(\sigma_0 + \sigma_1 y)^2} \, dx \right] \, dy} \\
  u(w) &= \frac{\int_{0}^{w} \exp \left[ -2 r_0 \frac{w}{\sigma_0^2} \right] \, dy}{\int_{0}^{w} \exp \left[ -2 r_0 \frac{w}{\sigma_0^2} \right] \, dy} \\
  u(w) &= \frac{1 - e^{-\frac{2r_0 w}{\sigma_0^2}}}{1 - e^{-\frac{2r_0 w}{\sigma_0^2}}} \\
  u(w) &= \frac{\int_{0}^{w} \exp \left[ \frac{1}{\sigma_1} \cdot \frac{2r_0}{\sigma_0 + \sigma_1 y} \right] \, dy}{\int_{0}^{w} \exp \left[ \frac{1}{\sigma_1} \cdot \frac{2r_0}{\sigma_0 + \sigma_1 y} \right] \, dy} \\
  u(w) &= \frac{\int_{0}^{w} \exp \left[ \frac{1}{\sigma_1} \cdot \frac{2r_0}{\sigma_0 + \sigma_1 y} \right] \, dy}{\int_{0}^{w} \exp \left[ \frac{1}{\sigma_1} \cdot \frac{2r_0}{\sigma_0 + \sigma_1 y} \right] \, dy}
\end{align*}
\]

Model 1  Model 2 (Numerical Integration)
Step 3 Continued…Solve the ODE in general

\[ u(w) = \frac{\int_0^w \exp \left[ -2 \int_0^y \frac{r_0 + r_1 x}{(\sigma_0 + \sigma_1 x)^2} \, dx \right] \, dy}{\int_0^w \exp \left[ -2 \int_0^y \frac{r_0 + r_1 x}{(\sigma_0 + \sigma_1 x)^2} \, dx \right] \, dy} \]

\[ u(w) = \frac{1 - e^{-2r_0 \sigma_0 w}}{1 - e^{-2r_0 \sigma_0 w}} \]

Model 1 \hspace{1cm} Model 2 (Numerical Integration) \hspace{1cm} Model 3 (A&S Approx)
Step 3 Continued... Solve the ODE in general

\[ u(w) = \frac{\int_0^w \exp \left[ -2 \int_0^y \frac{r_0 + r_1 x}{(\sigma_0 + \sigma_1 x)^2} \, dx \right] \, dy}{\int_0^w \exp \left[ -2 \int_0^y \frac{r_0 + r_1 x}{(\sigma_0 + \sigma_1 x)^2} \, dx \right] \, dy} \]

\[ u(w) = \frac{1 - e^{-\frac{2r_0}{\sigma_0}w}}{1 - e^{-\frac{2r_0}{\sigma_0}w_u}} \]

Model 1

Model 2 (Numerical Integration)

Model 3 (A&S Approx)

Model 4 (Numerical Integration)
Step 3 Continued…Solve the ODE in general

\[
\begin{align*}
    u(w) &= \frac{\int_0^w \exp \left[ -2 \int_0^y \frac{r_0 + r_1 x}{(\sigma_0 + \sigma_1 x)^2} \, dx \right] \, dy}{\int_0^w \exp \left[ -2 \int_0^y \frac{r_0 + r_1 x}{(\sigma_0 + \sigma_1 x)^2} \, dx \right] \, dy} \\
    u(w) &= 1 - e^{-\frac{2r_0}{\sigma_0^2} w} \\
    u(w) &= \int_0^w \exp \left[ \frac{1}{\sigma_1} \cdot \frac{2r_0}{\sigma_0 + \sigma_1 y} \right] \, dy \\
    u(w) &= \int_0^w \exp \left[ -\frac{r_1}{\sigma_0^2} y^2 \right] \, dy \\
    u(w) &= \int_0^w \exp \left[ -\frac{r_1}{\sigma_1^2} y^2 \right] \, dy \\
    u(w) &= \int_0^w \exp \left[ -\frac{r_1}{\sigma_0^2} y^2 \right] \, dy
\end{align*}
\]

Model 1

Model 2 (Numerical Integration)

Model 3 (A&S Approx)

Model 4 (Numerical Integration)

Model 5--The Full Monty
Step 4: Apply the MLE Parameters from the 5 Best Models and compute

\[ \bar{u}(w) = \sum_{i=1}^{5} AIC_w(i)u_i(w) \]

\[ \bar{u}^2(w) = \sum_{i=1}^{5} AIC_w(i)u_i^2(w) \]

\[ \text{Var}(u(w)) = \bar{u}^2(w) - (\bar{u}(w))^2 \]
Step 4: Apply the MLE Parameters from the 5 Best Models and compute

\[
\bar{u}(w) = \sum_{i=1}^{5} AIC_w(i)u_i(w)
\]

\[
\bar{u}^2(w) = \sum_{i=1}^{5} AIC_w(i)u_i^2(w)
\]

\[
\text{Var}(u(w)) = \bar{u}^2(w) - (\bar{u}(w))^2
\]
Step 4 Continued… The Role of Model Uncertainty
Step 5: Now consider parameter uncertainty

Use MCMC methods to obtain distributions for the parameters in each of the models. Estimate the joint posterior for the model parameters, denoted by \( \nu \).
Step 5: Now consider parameter uncertainty

Use MCMC methods to obtain distributions for the parameters in each of the models. Estimate the joint posterior for the model parameters, denoted by $\nu$

Then the posterior predictive distribution for the success probability $u(w)$ can be determined by applying the formula for $u(w)$ to draws of model parameters from their joint posterior distribution.
Step 5: Now consider parameter uncertainty

Use MCMC methods to obtain distributions for the parameters in each of the models. Estimate the joint posterior for the model parameters, denoted by $\nu$

Then the posterior predictive distribution for the success probability $u(w)$ can be determined by applying the formula for $u(w)$ to draws of model parameters from their joint posterior distribution.

$$p_1(\nu \mid W, dW) = \frac{p_0(\nu) L(dW, W \mid \nu)}{\int p_0(\nu) L(dW, W \mid \nu) d\nu} \propto p_0(\nu) L(dW, W \mid \nu)$$

Prior on parameters (uniform)  Gaussian likelihood of $dW$ given $W$ and parameters
Step 5 Continued…English Wikipedia, Model 1

**Posterior for r0**

**Posterior for s0**

**Model 1**

Density

$0.000 \quad 0.002 \quad 0.004 \quad 0.006$

$r0$

$0.000 \quad 0.004 \quad 0.008$

$s0$

$-200 \quad 0 \quad 100 \quad 200 \quad 300 \quad 400$

$100 \quad 200 \quad 300 \quad 400 \quad 500$

$-100 \quad 0 \quad 100 \quad 200 \quad 300 \quad 400$

$0 \quad 200 \quad 400 \quad 600 \quad 800 \quad 1000$

$0.0 \quad 0.2 \quad 0.4 \quad 0.6 \quad 0.8 \quad 1.0$
Step 5 Continued… English Wikipedia, Model 2
Step 5 Continued…English Wikipedia, Model 3
Step 5 Continued… English Wikipedia, Model 4
Step 5 Continued… English Wikipedia, Model 5
Step 5 Continued… We can do this for all 268 Wikipedias for which there are data (at, say, 18 months)
And Compare Predictions and Observations
And Compare Predictions and Observations

There is something else going on-
And Compare Predictions and Observations

There is something else going on-

- Saturation in small language Wikipedias (e.g. Koro, Matukar Panau)
And Compare Predictions and Observations

There is something else going on-

- Saturation in small language Wikipedias (e.g., Koro, Matukar Panau)
- Each language could have its own best model for the comparison
Outline

- Wikipedia -- a bit of history
- The population biology of competition between flour beetles
- The population biology of Wikipedians
- A few conclusions about the success and failures of social collaboration networks
- Advice to the young ones
Conclusions: Flour Beetles Help Illuminate the Wikipedia

- Identifying the local laws of global behavior of the Wikipedia?
Conclusions: Flour Beetles Help Illuminate the Wikipedia

• Identifying the local laws of global behavior of the Wikipedia?

• Explode the myth of exponential growth. Sigmoidal curves arise all the time in business (Sterman 2000) our work provides a variety of new insights about them.
Conclusions: Flour Beetles Help Illuminate the Wikipedia

- Identifying the local laws of global behavior of the Wikipedia?

- Explode the myth of exponential growth. Sigmoidal curves arise all the time in business (Sterman 2000) our work provides a variety of new insights about them.
Conclusions: Flour Beetles Help Illuminate the Wikipedia

• Identifying the local laws of global behavior of the Wikipedia?

• Explode the myth of exponential growth. Sigmoidal curves arise all the time in business (Sterman 2000) our work provides a variety of new insights about them.

Parsons’s Theorem

Birth and death
Or SDE with
Non-zero mean

Month since Dec 2000
Conclusions: Flour Beetles Help Illuminate the Wikipedia

- Identifying the local laws of global behavior of the Wikipedia?
- Explode the myth of exponential growth. Sigmoidal curves arise all the time in business (Sterman 2000) our work provides a variety of new insights about them.
Conclusions: Flour Beetles Help Illuminate the Wikipedia

• What are the local laws of global behavior of the Wikipedia?

• Explode the myth of exponential growth. Sigmoidal curves arise all the time in business (Sterman 2000) our work provides a variety of new insights about them.

![Graph showing the growth of editors over time with logistic growth and stochastic fluctuations highlighted.](image)

- **Birth and death**
- **Or SDE with Non-zero mean**
- **Logistic growth, stochastic effects can be ignored**
- **Diffusion-based fluctuations**

*Parsons’s Theorem*.

![Image of Parsons](image)
Conclusions: Flour Beetles Help Illuminate the Wikipedia

• Sustainability of Wikis

  – Important to focus on the size of the potential population
Conclusions: Flour Beetles Help Illuminate the Wikipedia

• Sustainability of Wikis
  
  – Important to focus on the size of the potential population
  
  - Are there ways of modifying r’s and σ’s to increase $u(w)$
Conclusions: Flour Beetles Help Illuminate the Wikipedia

• Sustainability of Wikis

  – Important to focus on the size of the potential population

  - Are there ways of modifying r’s and σ’s to increase \( u(w) \)

• Advice on Social Collaboration Networks

  - Can the early behavior lead to advice for investors and thus avoid the Concorde Fallacy (maybe another talk)?
Outline

• Wikipedia -- a bit of history

• The population biology of competition between flour beetles

• The population biology of Wikipedians

• A few conclusions about the success and failures of social collaboration networks

• Advice to the young ones
Lessons Learned in 40 Years of Mathematical Biology
Lessons Learned in 40 Years of Mathematical Biology

• The highest function of ecology is understanding consequences (Frank Herbert, *Dune*)
Lessons Learned in 40 Years of Mathematical Biology

• The highest function of ecology is understanding consequences (Frank Herbert, *Dune*)

• Variation is not noise. It is the core of biology.
Lessons Learned in 40 Years of Mathematical Biology

• The highest function of ecology is understanding consequences (Frank Herbert, *Dune*)

• Variation is not noise. It is the core of biology.

• The simpler the mathematical tool, the more likely it is to deliver the goods (John Hammersley)
Lessons Learned in 40 Years of Mathematical Biology

• The highest function of ecology is understanding consequences (Frank Herbert, *Dune*)

• Variation is not noise. It is the core of biology.

• The simpler the mathematical tool, the more likely it is to deliver the goods (John Hammersley)

• I would not give a fig for simplicity this side of complexity, but I would give my life for simplicity on the other side of complexity (Oliver Wendell Holmes)
And if You Don’t Care About Biology, the location of our future Nature and Technology ARE variable. AND much of that variability can be understood
One Final Acknowledgment: We Owe Our Teachers A Debt that We Can Never Repay to Them

Don Ludwig

Colin Clark

Ed Granirer

Lionel Harrison

Bob Snider

Jim Varah

Jim Zidek
wikipatterns

a practical guide to improving productivity and collaboration in your organization

Stewart Mader

Social selection and peer influence in an online social network

Kevin Lewis*,b,1, Marco Gonzalez*a, and Jason Kaufman*b

*Department of Sociology and "Berkman Center for Internet and Society, Harvard University, Cambridge, MA 02138; and "Behavioral Sciences Depart., Santa Rosa Junior College, Santa Rosa, CA 95401

Biologists using Social-Networking Sites to Boost Collaboration

Author(s) : Mark Crawford
Published By: American Institute of Biological Sciences