Modeling and Predicting the Growth and Death of Websites

Bruno Ribeiro

Carnegie Mellon University
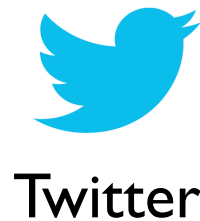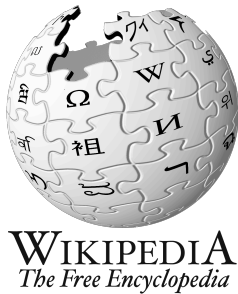
Rio de Janeiro, Mar 2014

Carnegie Mellon
School of Computer Science

Today's Information-Rich World

▸ WWW creates a wealth of information easily accessible

Wikipedia   Twitter   Blogs   Facebook

▸ But information wealth is dearth of something else
  ◦ If information consumes something it creates a resource allocation problem
  ◦ Wealth of information ➔ poverty of attention

Carnegie Mellon
School of Computer Science

Information Processing Systems

Many information systems are designed as if information was scarce

e.g.: Twitter   Facebook

when the problem is just the opposite.

"A wealth of information creates a poverty of attention"
- Herbert A. Simon, 1969

Herbert A. Simon and the Future of Computing

▸ Herbert A. Simon
  ◦ Turin Award Winner, 1975
  ◦ Nobel Laureate in Economics, 1978
  ◦ CMU faculty

Late 60's Simon's warning:
machines that "talk" more than "listen" and "think" exacerbate information overload

Simon's solution: Information processing systems should be designed to conserve attention

Twitter   Facebook

Carnegie Mellon
School of Computer Science

Today's Information Overload

Piano cats are **not** scarce

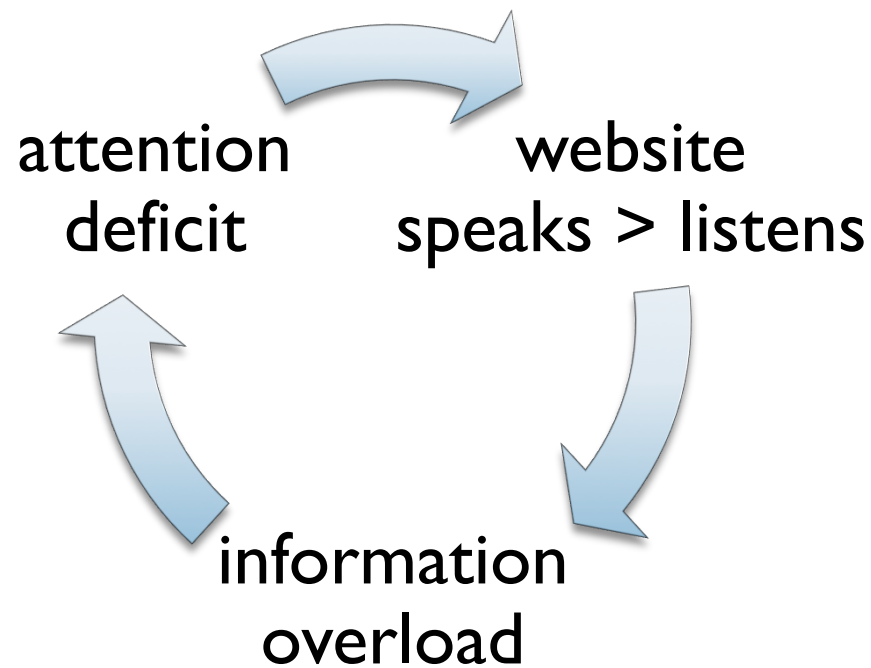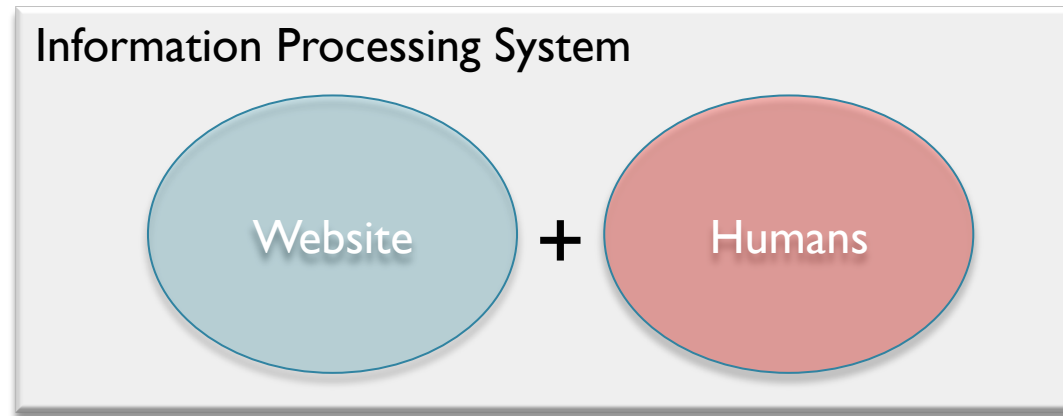They are over **abundant**,

    desperately fighting for our attention

Same with content on Facebook, TeaPartyNation.com, OccupyWallSt.org, Blogs, etc.

Carnegie Mellon
School of Computer Science

My View of Modern Information Systems

Information Processing System

Website + Humans

attention deficit → website speaks > listens → information overload → attention deficit

Bruno Ribeiro,   http://www.cs.cmu.edu/~ribeiro

Website Attention ⇔ Content Feedback Loop

▸ Facebook
attention
/content

Bruno Ribeiro
March 3

My dog refuses to come inside!

| 🏷 Tag Photo | 📍 Add Location | ✎ Edit |

Like · Comment · Stop Notifications · Share · Edit

Technology Adoptions
(Website popularity)

Carnegie Mellon
School of Computer Science

Outline

▸ Existing theory of adoptions

▸ Key to success lies in failures
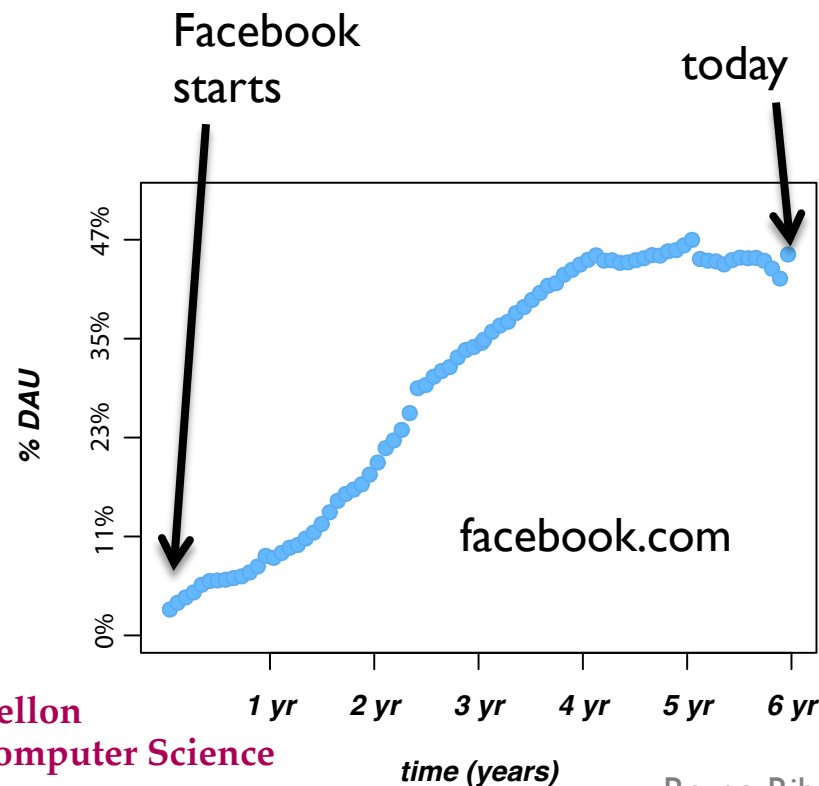
▸ Model

▸ Model predictions

▸ Discussion

Carnegie Mellon
School of Computer Science

Outline

▸ **Existing theory of adoptions**

▸ Key to success lies in failures

▸ Model

▸ Model predictions

▸ Discussion

Carnegie Mellon
School of Computer Science

Preliminaries: Internet success stories

▸ Measuring Internet startup value is hard, open problem (Cauwels & Sornette '12)

▸ widely used metric of success:
Daily fraction of Active Users (%DAU) – as % of active Internet pop.

Facebook starts

today

facebook.com

% DAU

47%

35%

23%

11%

0%

1 yr   2 yr   3 yr   4 yr   5 yr   6 yr

time (years)

**Carnegie Mellon
School of Computer Science**

data from Alexa.com

Adoption of Innovations

How do people adopt new behavior or technology?

Mansfield'61, Rogers'03

fraction of adopters (%)

rate of adoption

C

C/2

0
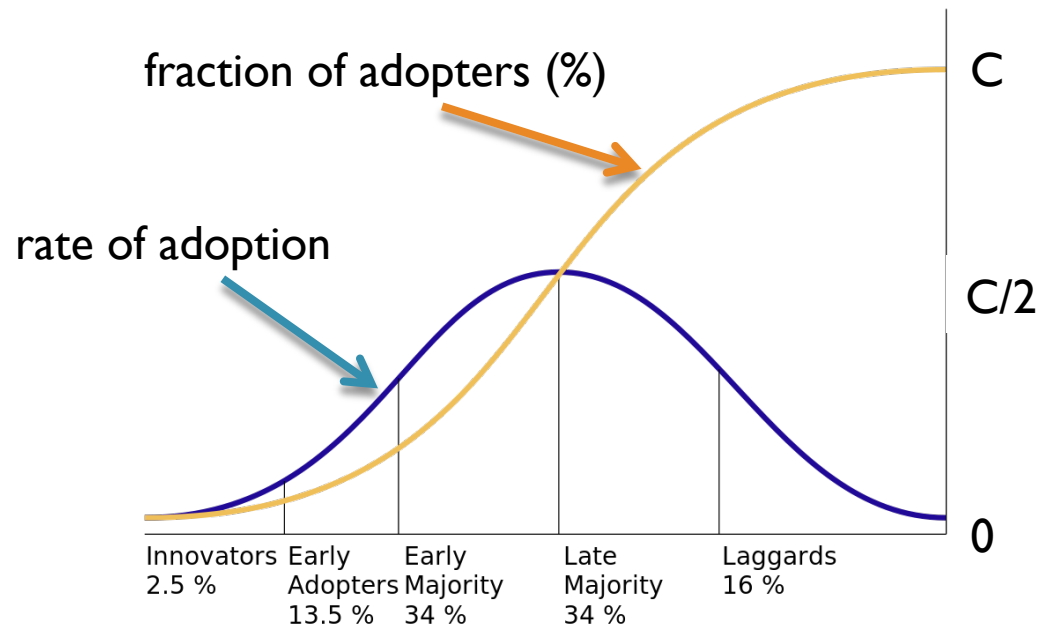
C - fraction of population
   that adopts

some will
never adopt

| Innovators 2.5 % | Early Adopters 13.5 % | Early Majority 34 % | Late Majority 34 % | Laggards 16 % |

Carnegie Mellon
School of Computer Science

Bruno Ribeiro,   http://www.cs.cmu.edu/~ribeiro

sources: wikipedia, arstechna, eideard.com

Membership Growth in Successful Websites

Rate of growth shows
adoption of innovations
signature

new subscribers/yr
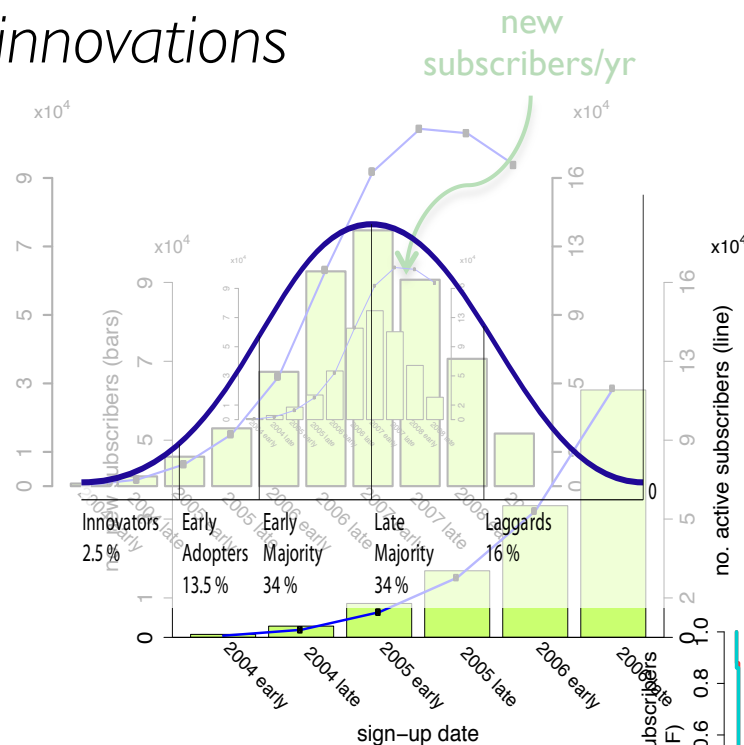
MySpace.com

data collected in Feb. 2009

little Facebook influence in 2007

bars – new members (adoptions)
line – active members

Carnegie Mellon
School of Computer Science

Bruno Ribeiro,   http://www.cs.cmu.edu/~ribeiro

Vast Adoption Literature

Economics:
- (Mansfield '63)
- (Katz&Shapiro'85)
- (Farrell&Saloner '86)
- (Choi '94)
- (Arthur '94)

Marketing:
- (Bass '69)
- (Fisher&Pry '71)

Computer Science:
- (Kempe et al '03)
- (Zhao et al., IMC'12)
- (Leskovec et al., SIGKDD'08)
- (Ugander et al., PNAS'12)
- (Aral&Walker, Science'12)

Sociology:
- (Ryan&Gross'49)
- (Everett '62, '03)
- (Rogers '03)
- (Centola '12)

Carnegie Mellon
School of Computer Science

Bruno Ribeiro,   http://www.cs.cmu.edu/~ribeiro

And Popularity Growth of Successful Websites *Seems* to Follow Adoption Models

▸ % DAU of successful websites
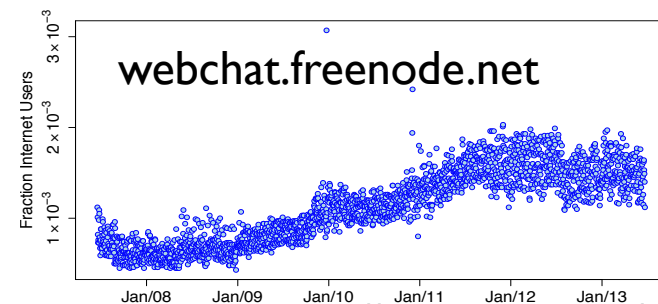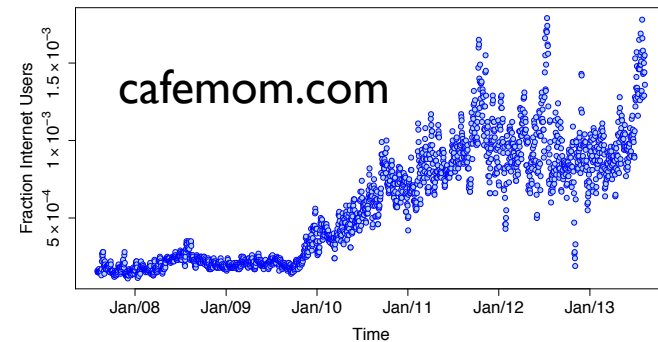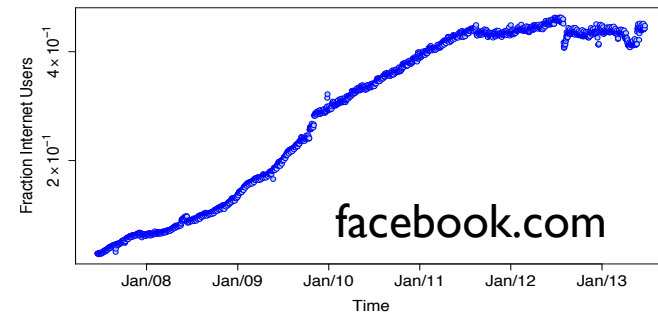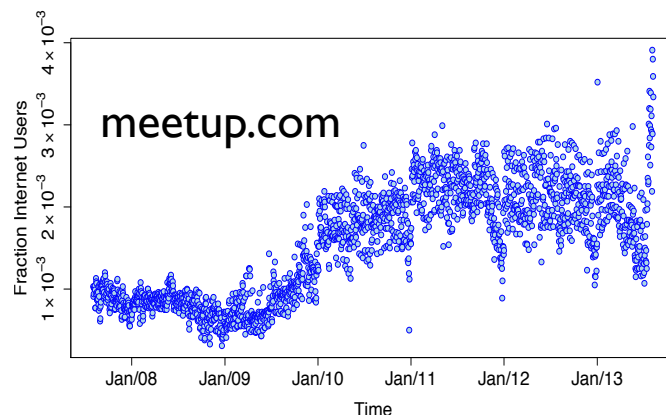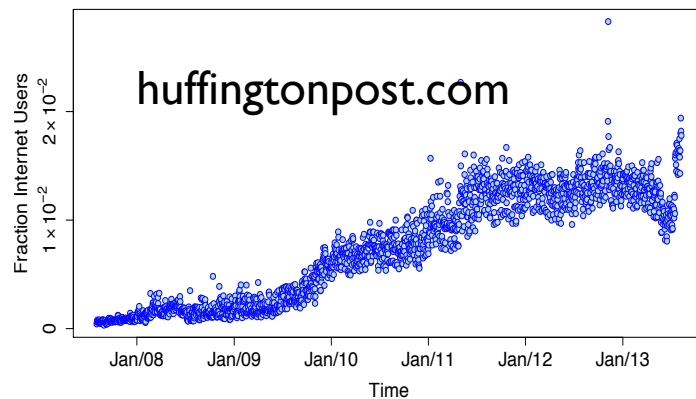
huffingtonpost.com

facebook.com

meetup.com

cafemom.com

webchat.freenode.net

Outline

▸ Existing theory of adoptions

▸ Key to success lies in failures

▸ Model

▸ Model predictions

▸ Discussion

Carnegie Mellon
School of Computer Science

While Failed Websites

Some startups are short-lived fads
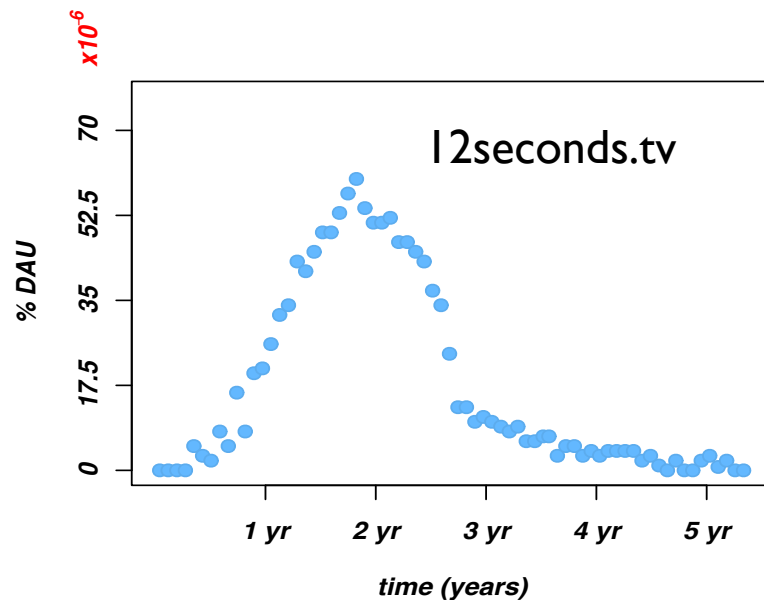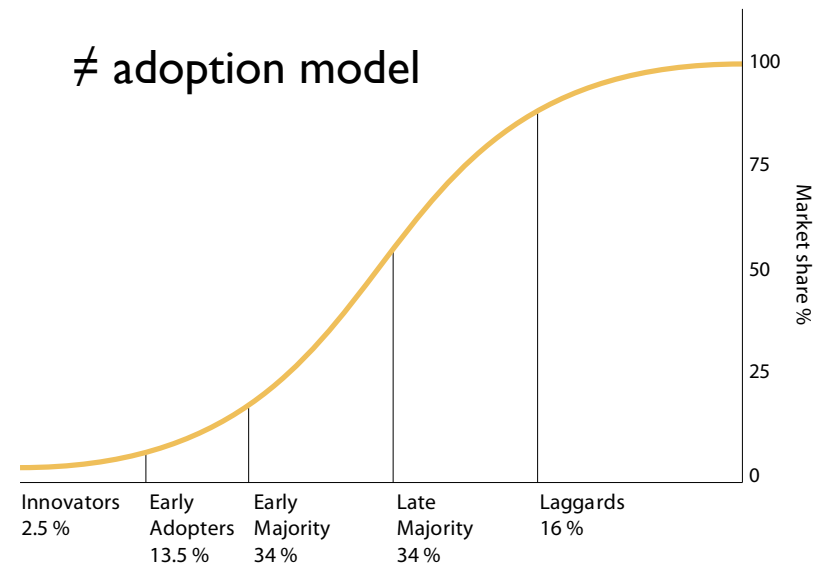
Literature dismiss failures!
"died from competition"
"died of incompetence"

Bruno Ribeiro, http://www.cs.cmu.edu/~ribeiro

Carnegie Mellon
School of Computer Science

Can Attention Scarcity Explain These Failures?

Bruno Ribeiro, Modeling and Predicting the Growth and Death of Membership-based Websites, WWW 2014

Media coverage:

Bruno Ribeiro,   http://www.cs.cmu.edu/~ribeiro

Outline

▸ Existing theory of adoptions

▸ Key to success lies in failures

▸ **Model**

▸ Model predictions

▸ Discussion

Carnegie Mellon
School of Computer Science

Positive & Negative Attention Loops
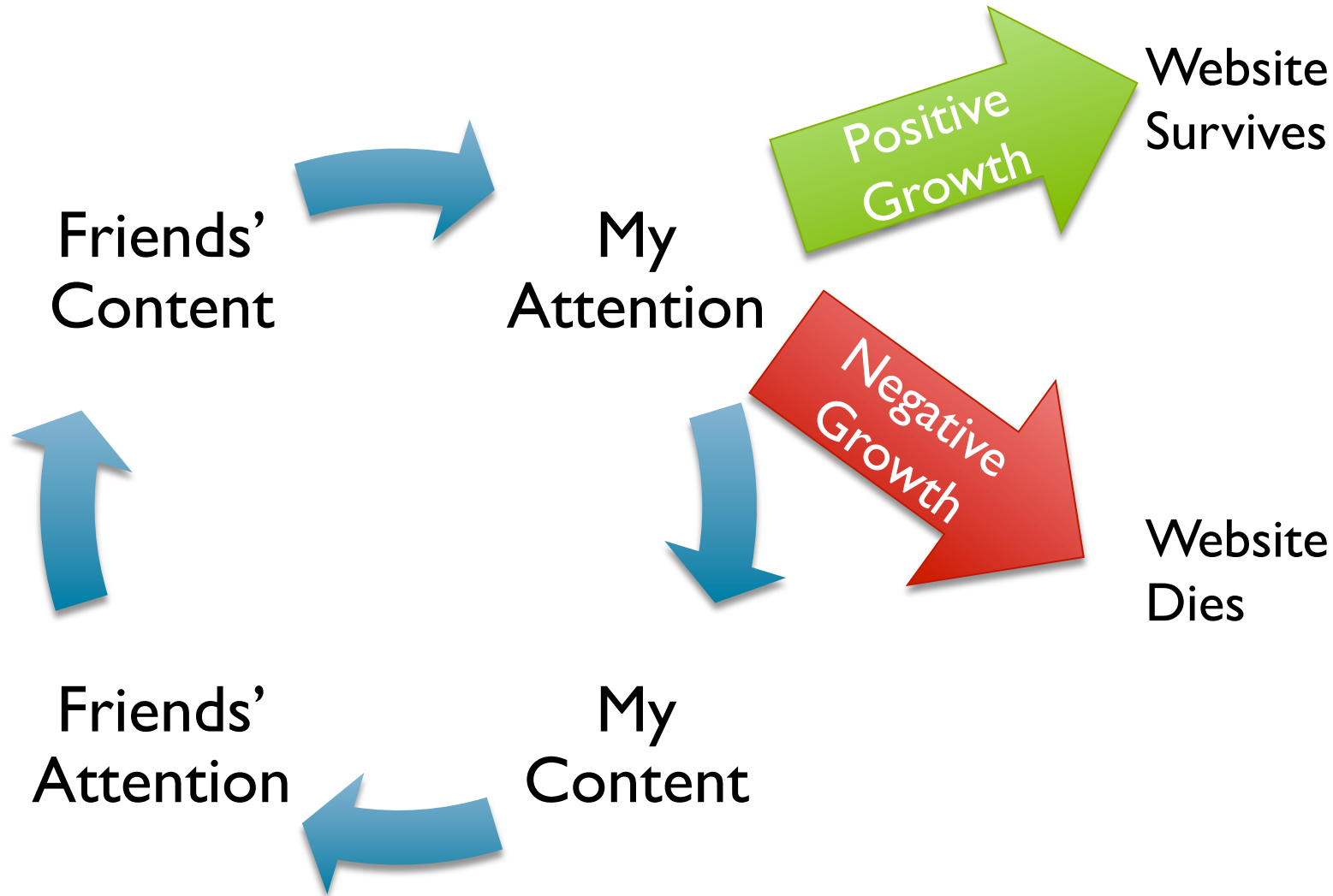
Mathematics

Population-level model

Carrying capacity parameter

- C — fraction of active Internet population that will eventually become members

Carnegie Mellon
School of Computer Science

Attention-Seeking Mechanics (activity - inactivity)

Modeled as reaction-decay process

Reaction:
$$A + I \xrightarrow{\alpha} 2A$$

active user   inactive user

Decay:
$$A \xrightarrow{\beta} I$$

active

active

network effect

loses interest

inactive

Carnegie Mellon
School of Computer Science

22
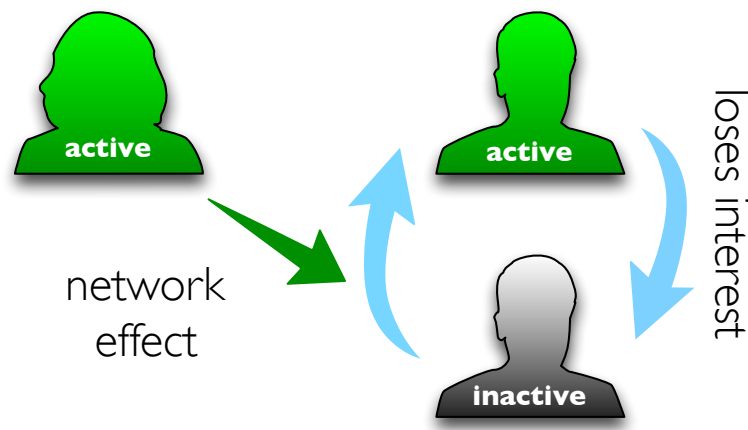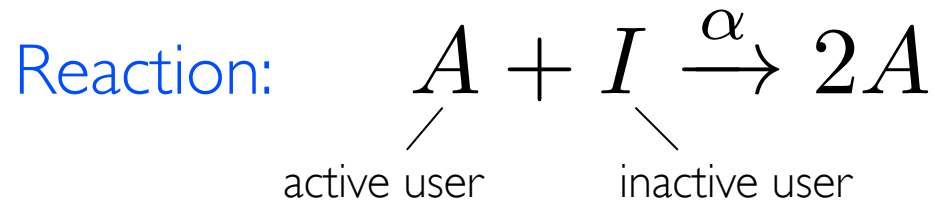
Adoptions Reaction-Diffusion Processes

Modeled as reaction-diffusion process

Media & Marketing diffusion: $\quad U \xrightarrow{\lambda} A$

Word-of-mouth reaction: $\quad A + U \xrightarrow{\gamma} 2A$

brings new user   brings new user

active   active

loses interest

network effect

inactive

Bruno Ribeiro,   http://www.cs.cmu.edu/~ribeiro
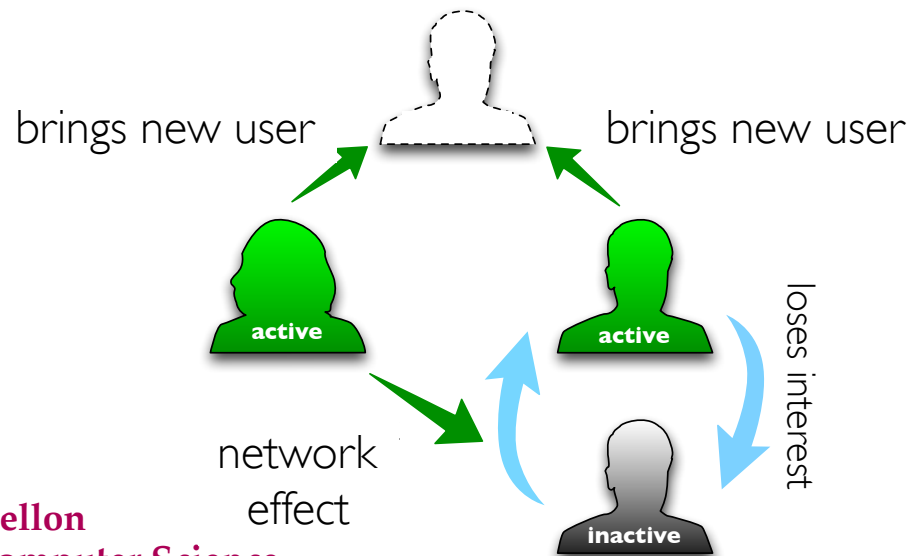
Complete Model

Variables:

- A(t) = % DAU at time t
- I(t) = % active Internet pop. of inactive members

Reaction-diffusion-decay differential equations:

$$\frac{dA(t)}{dt} = -\frac{1}{C}A^2(t)\gamma + \frac{1}{C}I(t)A(t)(\alpha - \gamma) + C\lambda$$
$$- A(t)(\beta + \lambda - \gamma) - I(t)\lambda\,,$$
$$\frac{dI(t)}{dt} = A(t)\beta - \frac{1}{C}I(t)A(t)\alpha\,,$$

Carnegie Mellon
School of Computer Science

Outline

▸ Existing theory of adoptions

▸ Key to success lies in failures

▸ Model

▸ **Model predictions**

▸ Discussion

Carnegie Mellon
School of Computer Science

Long-term Popularity Sustainability (signatures)

▸ Model predictions – long-term DAU (t' = t – t₀)

$$A(t) \approx \frac{(\beta - \alpha)e^{-(\beta-\alpha)t'}}{1 - \frac{\alpha}{C}e^{-(\beta-\alpha)t'}},$$

(a) self-sustaining — $\frac{\beta}{\alpha} < 1$, DAU (A(t)) vs t

(b) unsustainable — $\frac{\beta}{\alpha} \geq 1$, DAU (A(t)) vs t

Bruno Ribeiro,   http://www.cs.cmu.edu/~ribeiro
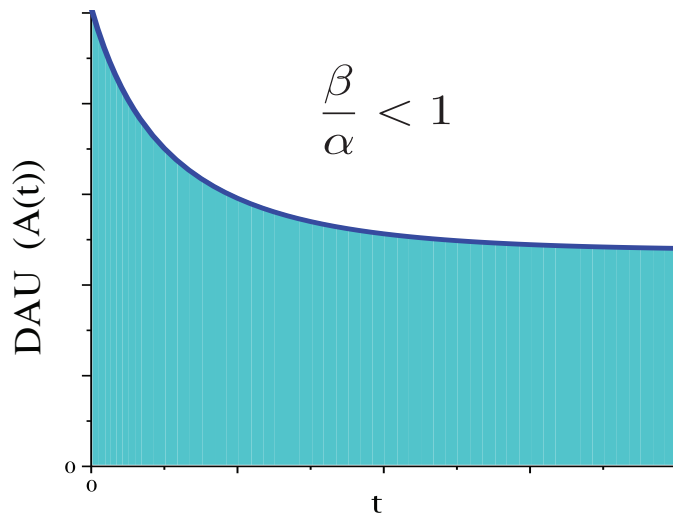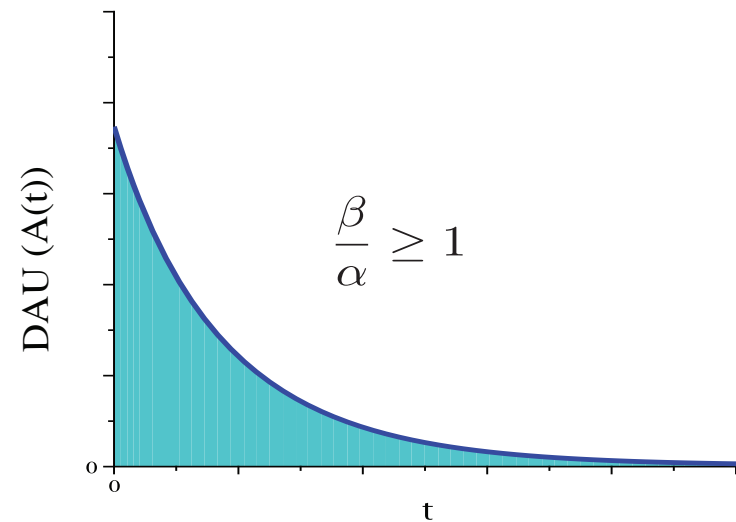
Signatures of Popularity Growth

▸ Model prediction: DAU signatures of growth
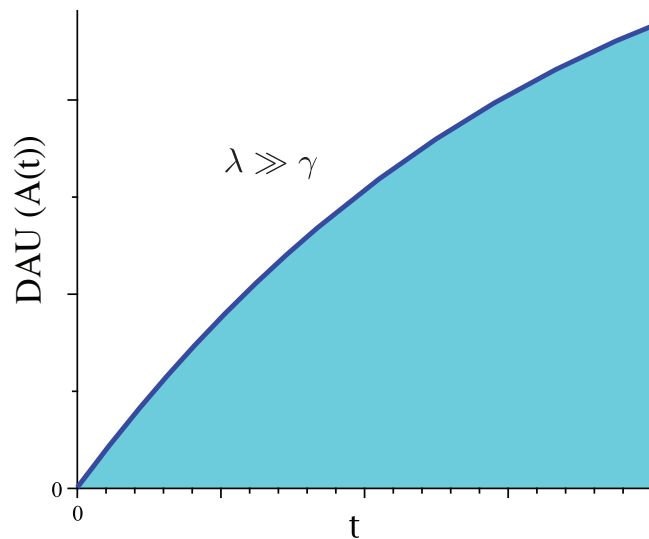
(a) media & marketing intensive
initial growth signature
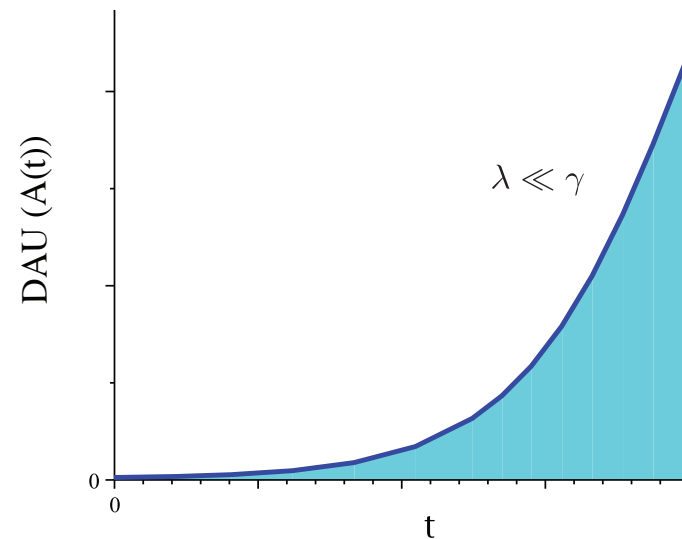
(b) word-of-mouth intensive
initial growth signature

Carnegie Mellon
School of Computer Science

Bruno Ribeiro,   http://www.cs.cmu.edu/~ribeiro

Datasets

▸ 22 websites

▸ from Alexa.com

▸ up to 6 years of DAU
(June 2007 – Sept 2013)

▸ **12seconds.tv:** "12seconds.tv is a Twitter-like video status service. It gives you 12 seconds to share video moments from your life" [15].

▸ **adaptu.com:** Membership-based online mobile wallet.

▸ **ashleymadison.com:** "Ashley Madison is a Married Dating service and social network for those engaged in relationships but looking to have an affair" [15].

▸ **brandstack.com:** "Brandstack lets designers create their own studios and sell design work to their peers or to directly consumers". It was sold in December 2011 to DesignCrowd [15].

▸ **cafemom.com:** "CafeMom is a social network site for moms, reaching an audience of more than 20 million users." [15].

▸ **community.babycenter.com:** The Community Baby Center, launched in 2008, is a social network for parents with young children of all ages to share their experiences.

▸ **facebook.com:** Online social network website.

▸ **flixter.com:** Flixster is a social networking site for movie fans.

▸ **formspring.me:** "Formspring helps people find out more about each other by sharing interesting & personal responses" [15].

▸ **huffingtonpost.com:** The Huffington Post is a leading left-leaning online news aggregator and producer.

▸ **linkedin.com:** Is a fast-growing online professional networking website.

▸ **marriedsecrets.com:** Married dating service and social network.

▸ **meetup.com:** Is a local community social network.

▸ **netflix.com:** Membership-based movie rental website.

▸ **occupywallst.org:** Is a website for people interested in the Occupy Wall Street movement.

▸ **patientslikeme.com:** Patientslikeme is a social networking site that allows people with similar diseases to share their experiences about treatments, doctors, and seek emotional support.

▸ **ruelala.com:** Membership-based online retail store.

▸ **teapartynation.com & teapartypatriots.org:** TeaPartyNation.com and TeaPartyPatriots.org are the official website of conservatives American political organizations considered part of the Tea Party movement.

▸ **true.com:** True is an online dating service that was founded in 2003 [15].

▸ **theblaze.com:** The Blaze is a conservative news and opinion website run by conservative pundit Glenn Beck [47].

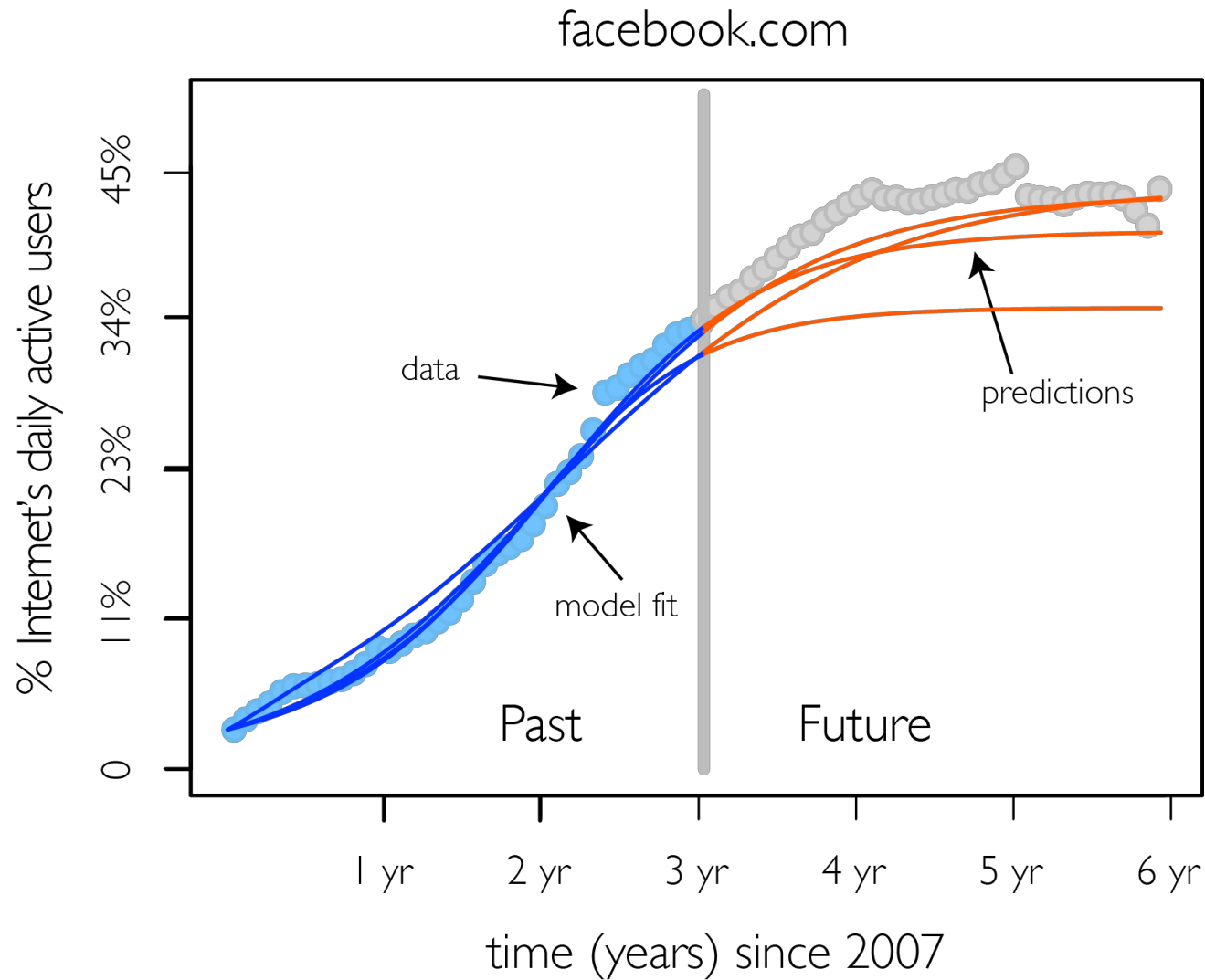▸ **webchat.freenode.net:** Is a web-based IRC chat server.

Carnegie Mellon
School of Computer Science
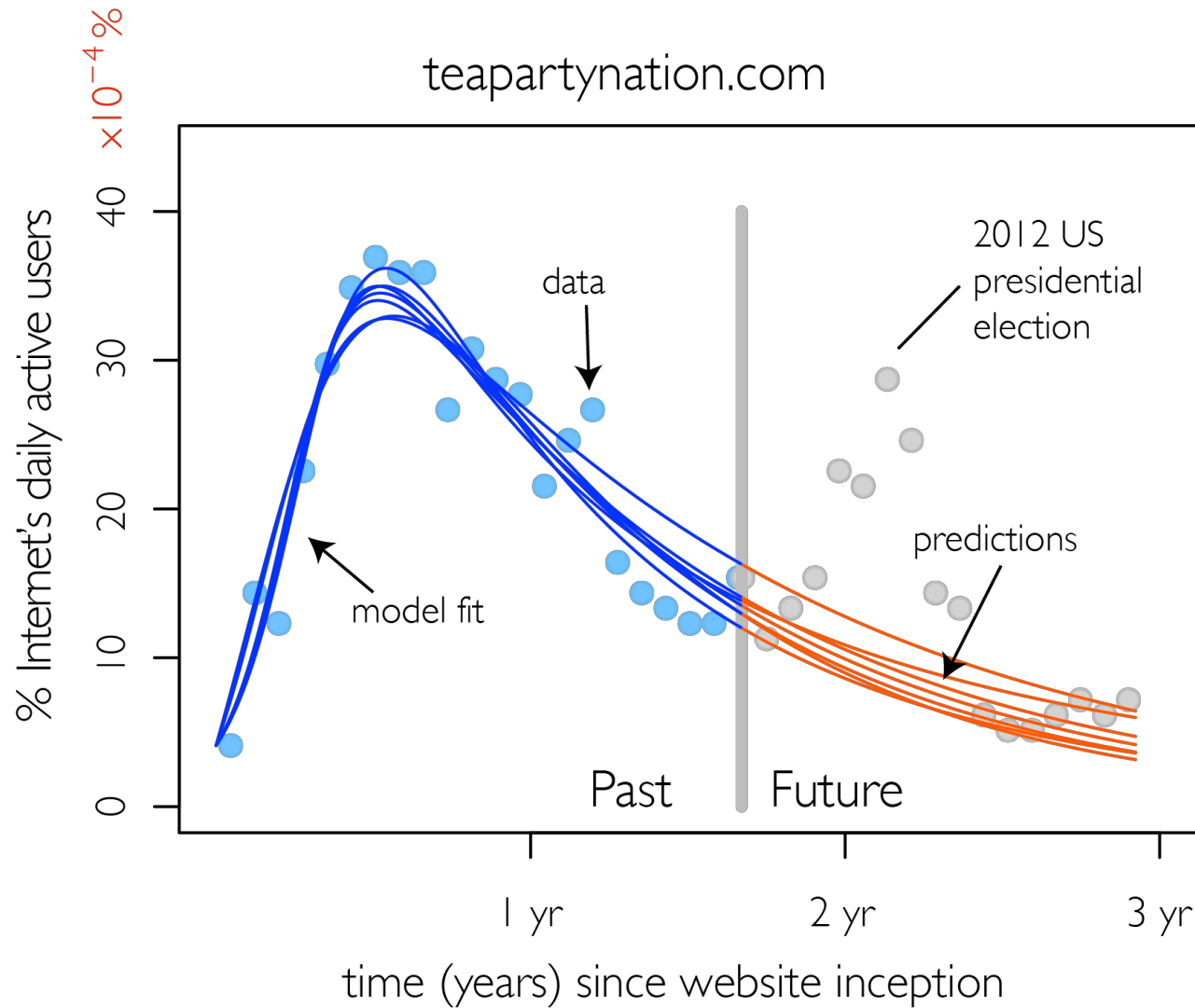
Automatic parameter fit
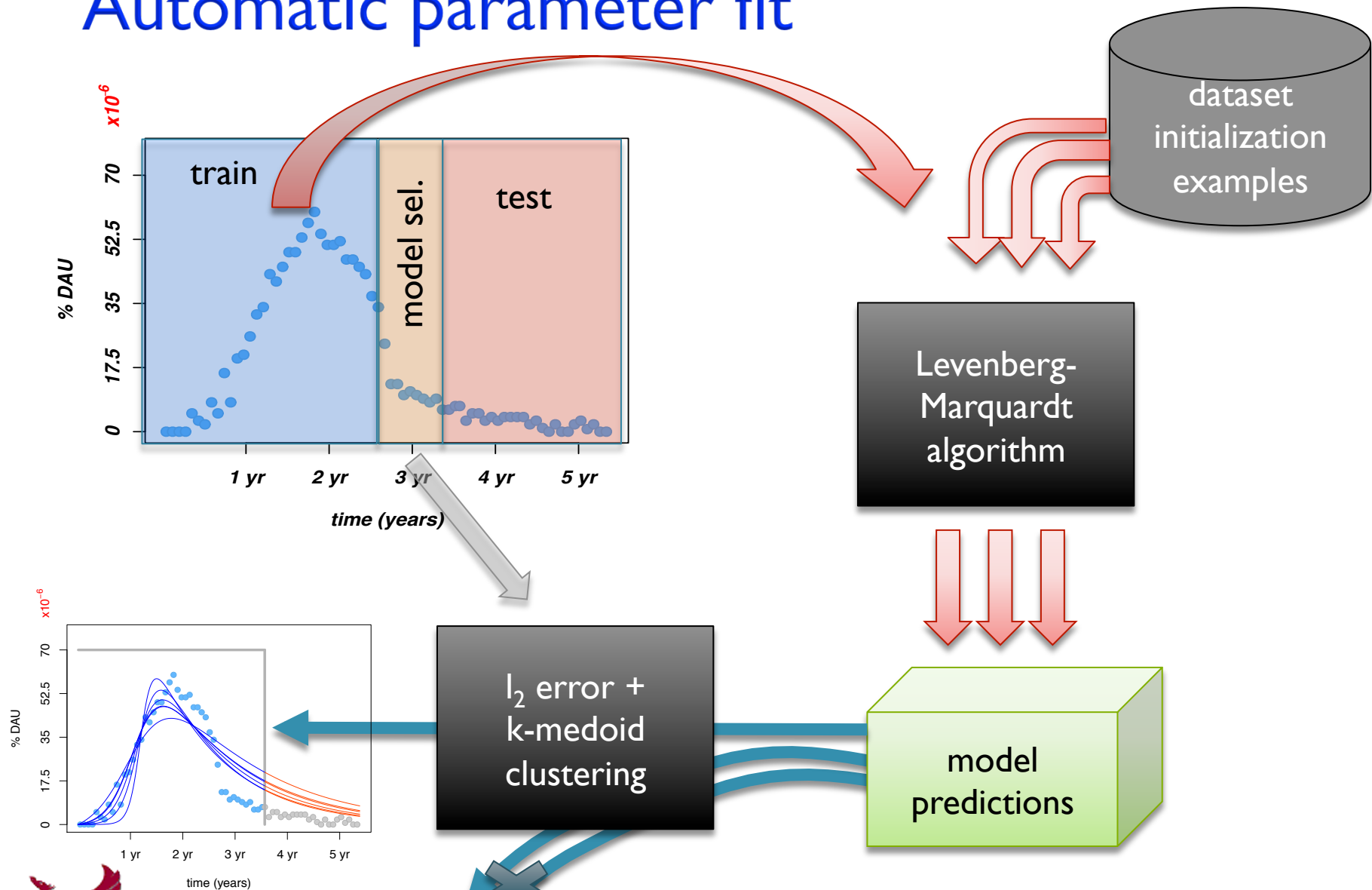
Predictions

Predictions for Facebook

facebook.com

Past

Future

% Internet's daily active users

data

model fit

predictions

time (years) since 2007

Predictions for TeaPartyNation.com

teapartynation.com

% Internet's daily active users — ×10^{-4} %

- data
- model fit
- 2012 US presidential election
- predictions

Past | Future

1 yr   2 yr   3 yr

time (years) since website inception

Predicted as unsustainable websites
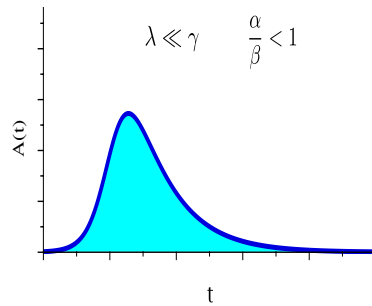
(a) Common unsustainable + word-of-mouth DAU shape.
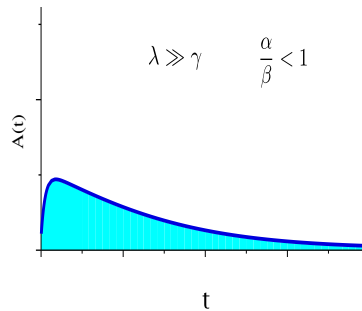
(b) Common unsustainable + media & marketing exposure DAU shape.
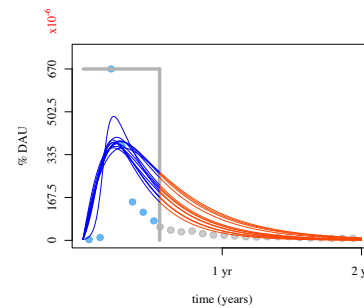
(c) occupywallst.org
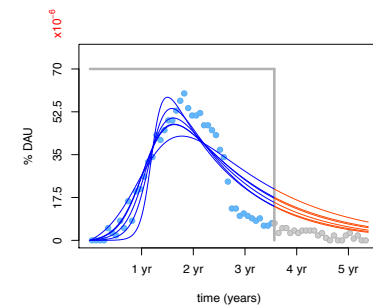
(d) 12seconds.tv
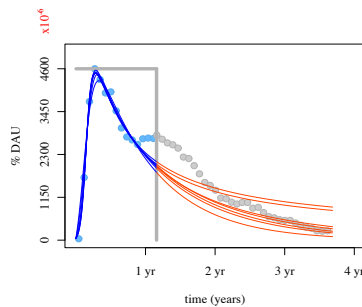
(e) formspring.me
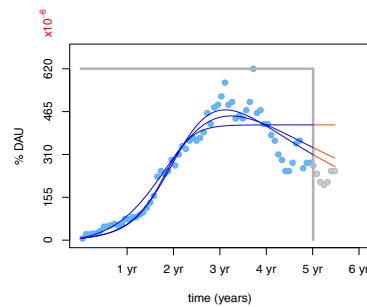
(f) ruelala.com
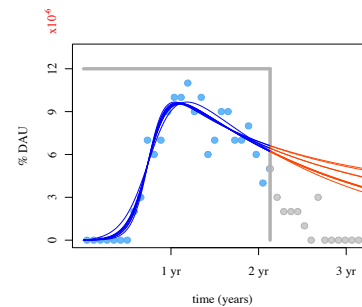
(g) adaptu.com
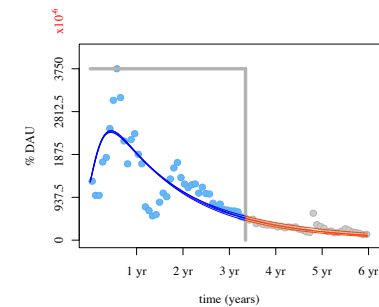
(h) flixster.com
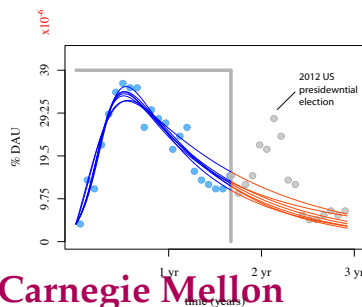
(i) teapartypatriots.com
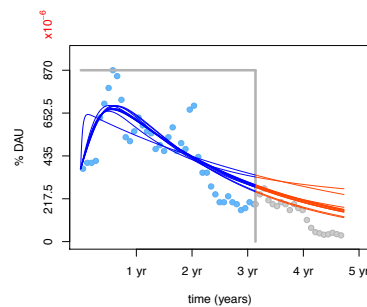
(j) true.com
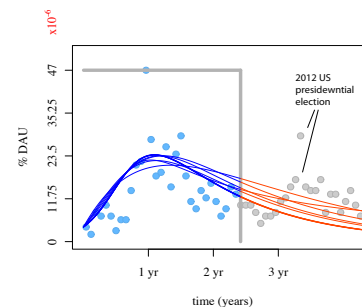
(k) teapartypatriots.org
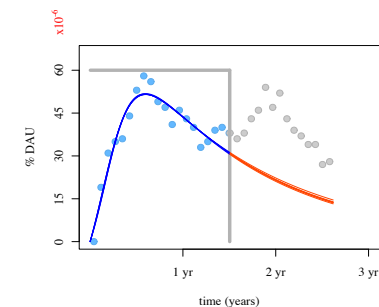
(l) brandstack.com (before acquisition)

Predicted as self-sustaining websites
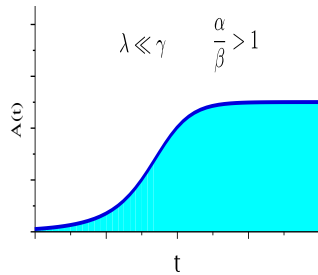
(a) Common self-sustaining + word-of-mouth DAU shape.
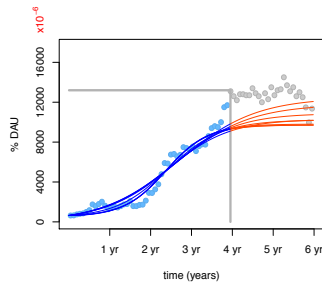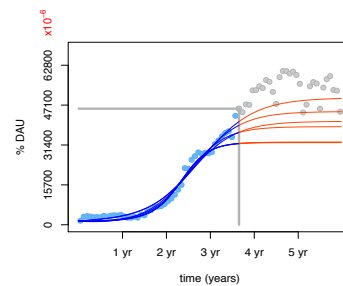
(b) huffingtonpost.com
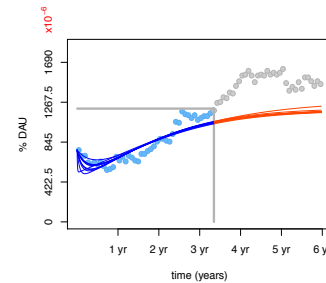
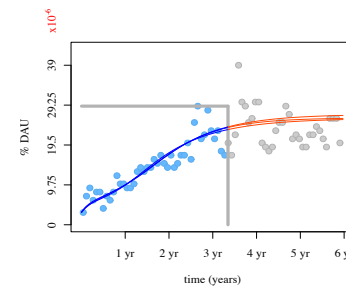(c) community.babycenter.com
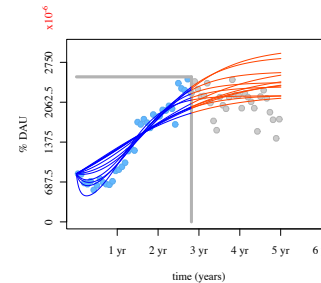
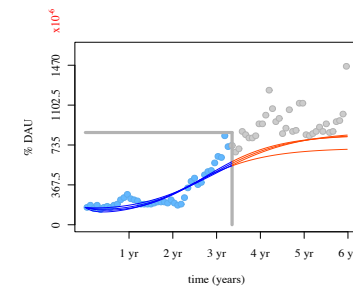(d) meetup.com
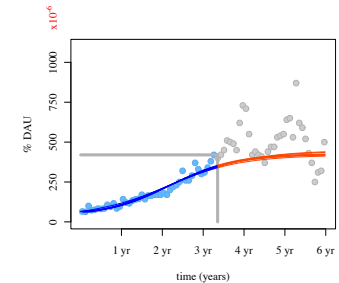
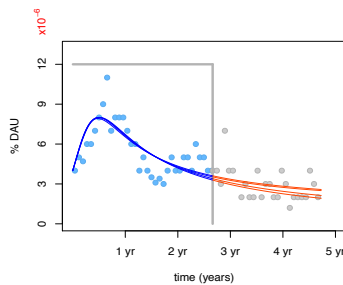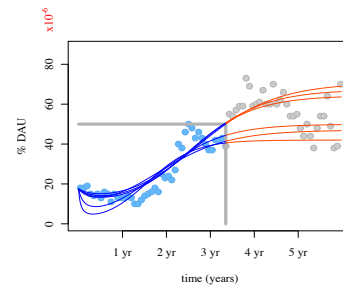(e) ashleymadison.com
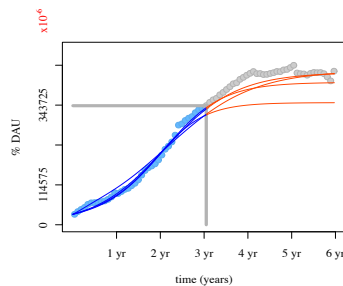
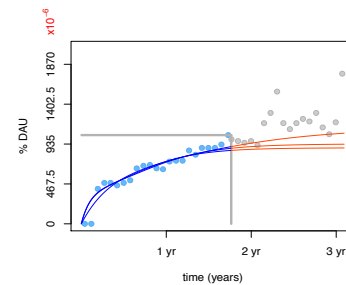(f) facebook.com
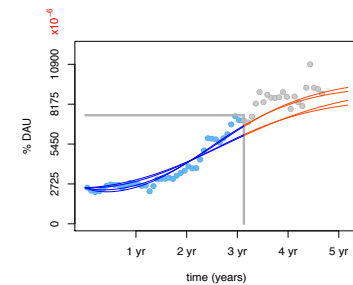
(g) linkedin.com

(h) patientslikeme.com

(i) cafemom.com

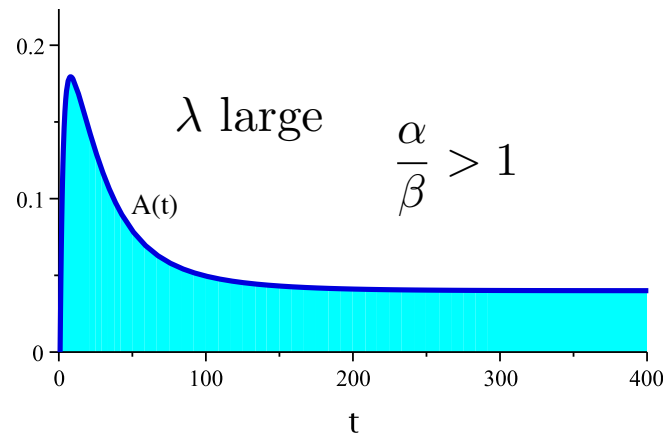(j) marriedsecrets.com

(k) webchat.freenode.net

(l) theblaze.com

(m) netflix.com

Carnegie Mellon
School of Computer Science

Bruno Ribeiro, http://www.cs.cmu.edu/~ribeiro

Another predicted behavior

λ large $\qquad \dfrac{\alpha}{\beta} > 1$

A(t)

A good model predicts the existence
of things not-yet-observed

Predicted behavior recently found!

media & marketing intensive
+
self-sustaining

λ large $\qquad \dfrac{\alpha}{\beta} > 1$
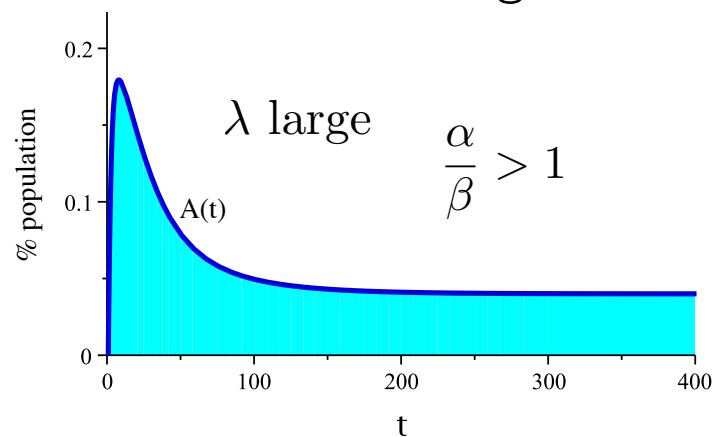
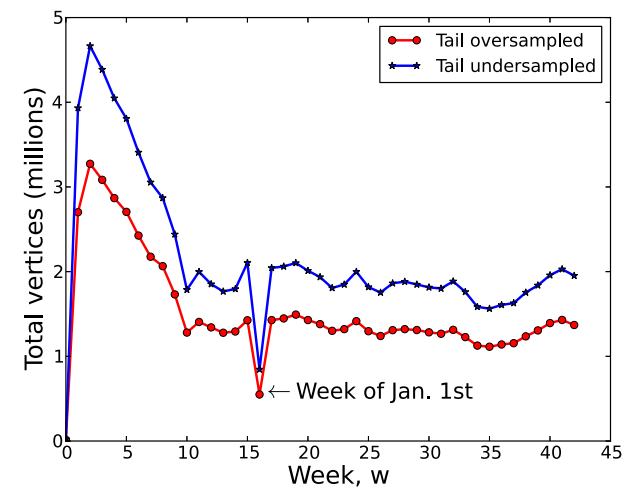…until Merrit & Clauset found after paper was on arXiv

Halo Reach massive multiplayer online game network

Merrit & Clauset (2013)

Carnegie Mellon
School of Computer Science

Bruno Ribeiro,   http://www.cs.cmu.edu/~ribeiro

Model Without Attention Feedback is Wrong

Without **Attention Feedback** websites never die
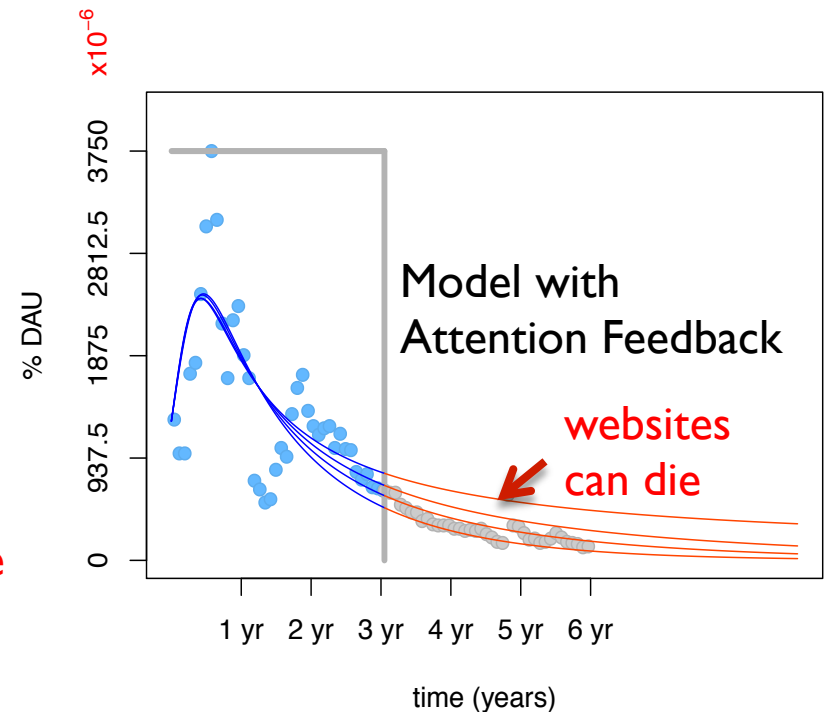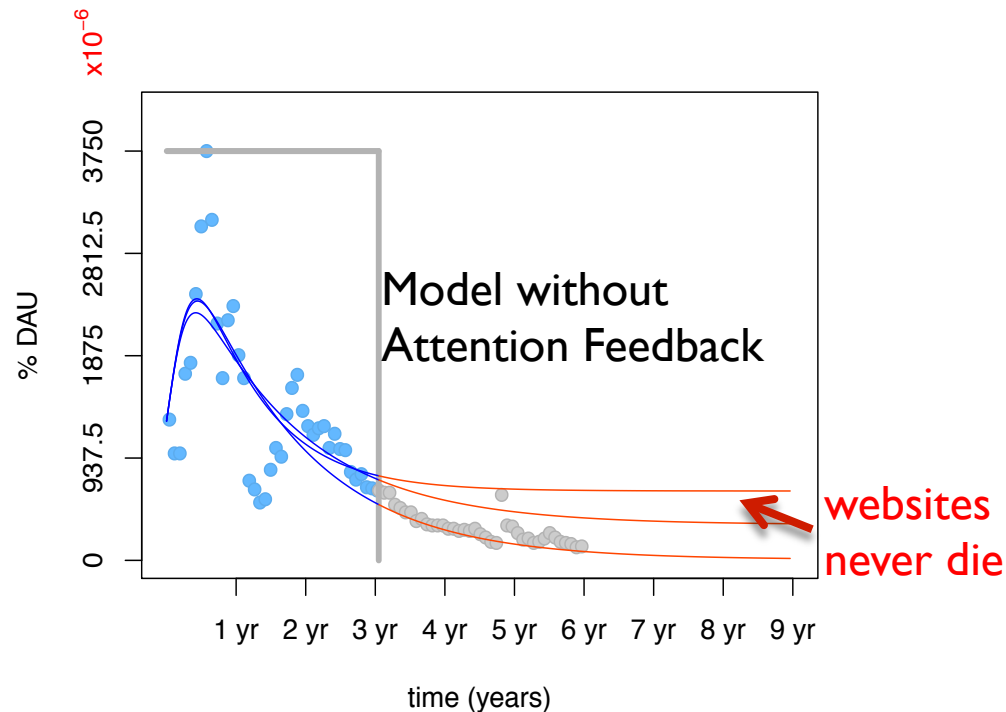
Parting Thoughts

Bruno Ribeiro,   http://www.cs.cmu.edu/~ribeiro

Parting Thoughts

▸ Facebook, Twitter, the Tea Party and Occupy Wall Street **rely on users** intensely generating content to **compete for attention**
  ◦ How modern Information Systems compete for popularity

▸ Model point to an even more sensationalist and competitive future for social media
  ◦ Website-addicted teens good for business
  ◦ Help users to create more **attention-grabbing** content

▸ Models of attention-activity dynamics may help us finally find a way to reduce our ever-increasing information overload

Carnegie Mellon
School of Computer Science
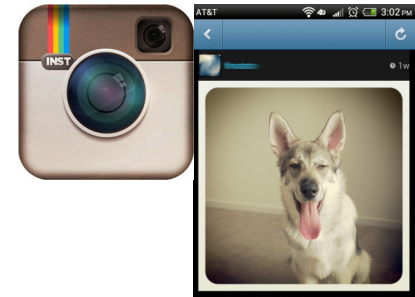
Thank you!

Bruno Ribeiro
ribeiro@cs.cmu.edu
http://www.cs.cmu.edu/~ribeiro/

A link to all data collection scripts and model-related code is provided in the paper
Bruno Ribeiro, **Modeling and Predicting the Growth and Death of Membership-based Websites**, *WWW 2014*

Carnegie Mellon
School of Computer Science