Information Diffusion

David Liben-Nowell

dlibenno@carleton.edu

Carleton College, Department of Computer Science

Portions are joint work with

Jon Kleinberg and Flavio Chierichetti

Central question:

How do ideas spread?

Ideas, disease, innovation, jokes spread continually via the global social network.

From: "[redacted]
To: "[redacted]
Subject: Fw: Hey Ladies! Make some noise -- please read
Date: Tue, 9 Sep 2008 15:20:15 -0700

----- Original Message -----
From: [redacted]
To: [redacted]
Sent: Tuesday, September 09, 2008 11:05 AM
Subject: FW: Hey Ladies! Make some noise -- please read

Looks like a constructive way to vent

On Tue, 9/9/08, [redacted] wrote:

> From: [redacted]
> Subject: Fwd: Hey Ladies! Make some noise -- please read
> To: [redacted]
> Date: Tuesday, September 9, 2008, 7:32 AM

> >From: "[redacted]
> >To: "[redacted]
> >Date: September 08, 2008 01:36:57 PM EDT
> >Subject: Fwd: Hey Ladies! Make some noise -- please read

We are writing to you because of the fury and dread we have felt since the announcement of Sarah Palin as the Vice-Presidential candidate for the Republican Party. We believe that this terrible decision has surpassed mere partisanship, and that it is a dangerous farce on the part of a pandering and rudderless Presidential candidate that has a real possibilty of becoming fact.

to [redacted] show details 7/2/05 Reply

this is actually kind of cool . . .

[redacted]
mobile [redacted]
fax [redacted]
email [redacted]
web [redacted]

-----Original Message-----
From: [redacted]
Sent: Friday, July 01, 2005 11:22 PM
To: [redacted]
Subject: Fw: (no subject)

----- Original Message -----
From: [redacted]
To: [redacted]
Sent: Tuesday, June 28, 2005 9:25 PM
Subject: (no subject)

I cdnuolt blveiee taht I cluod aulaclty uesdnatnrd waht I was rdanieg. The phaonmneal pweor of the hmuan mnid aoccdrnig to rscheearch taem at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the

believes in health care but is also super excited that she got the job!!!!!!!!!
11 minutes ago · Comment · Like

and am especially glad to have a job as my health benefits just ran out.
8 minutes ago

Congrats!
8 minutes ago

Congratulations! You go woman!
2 minutes ago

Write a comment...

No one should die because they cannot afford health care. No one should go broke because they get sick. No one should be denied medical care by their own insurer. If you agree, post as your status for today.
27 minutes ago · Comment · Like

believes that no one should die because they cannot afford health care, and no one should go broke because they get sick. If you agree, please post this as your status for the rest of the day.
28 minutes ago · Comment · Like

likes this.

Write a comment...

No one should die because they cannot afford health care, and no one should go broke because they get sick. If you agree, please post this as your status for the rest of the day.
35 minutes ago · Comment · Like

thinks that no one should die because they cannot afford health care, and no one should go broke because they get sick. If you agree, please post this as your status for the rest of the day.
42 minutes ago · Comment · Like

Central question:

How do ideas spread?

Ideas, disease, innovation, jokes spread continually via the global social network.

But *how?*

A known but typically unobserved process. How can we observe it?

A known but typically unobserved process. How can we observe it?

Pacific bluefin tuna transport Fukushima-derived radionuclides from Japan to California

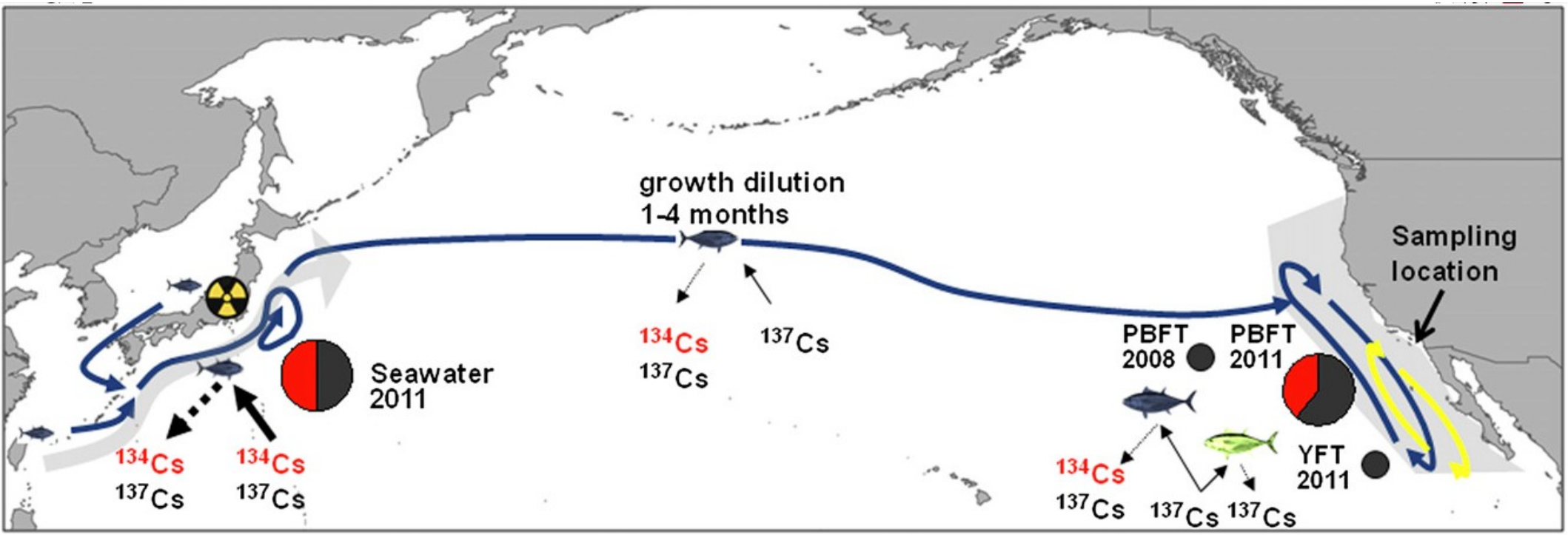
Daniel J. Madigan[a,1], Zofia Baumann[b], and Nicholas S. Fisher[b]

[a]Hopkins Marine Station, Stanford University, Pacific Grove, CA 93950; and [b]School of Marine and Atmospheric Sciences, Stony Brook University, Stony B
NY 11794

Edited by Karl K. Turekian, Yale University, North Haven, CT, and approved April 25, 2012 (received for review March 22, 2012)

The Fukushima Dai-ichi release of radionuclides into ocean waters caused significant local and global concern regarding the spread of radioactive material. We report unequivocal evidence that Pacific bluefin tuna, *Thunnus orientalis*, transported Fukushima-derived radionuclides across the entire North Pacific Ocean. We measured γ-emitting radionuclides in California-caught tunas and found ^{134}Cs (4.0 ± 1.4 Bq kg^{-1}) and elevated ^{137}Cs (6.3 ± 1.5 Bq kg^{-1}) in 15 Pacific

in their first year or early in their second (5). Thus, all blu
between years 1–2 (here, 2-y-old PBFT) caught during summ
the eastern Pacific must have migrated from the western Pa
within several months of capture. Waters north of the Kuro
Current (Fig. 1*A*) showed high radionuclide concentration
spring 2011 (3), and juveniles make extensive use of this re
before their eastward migration to the CCLME (6).



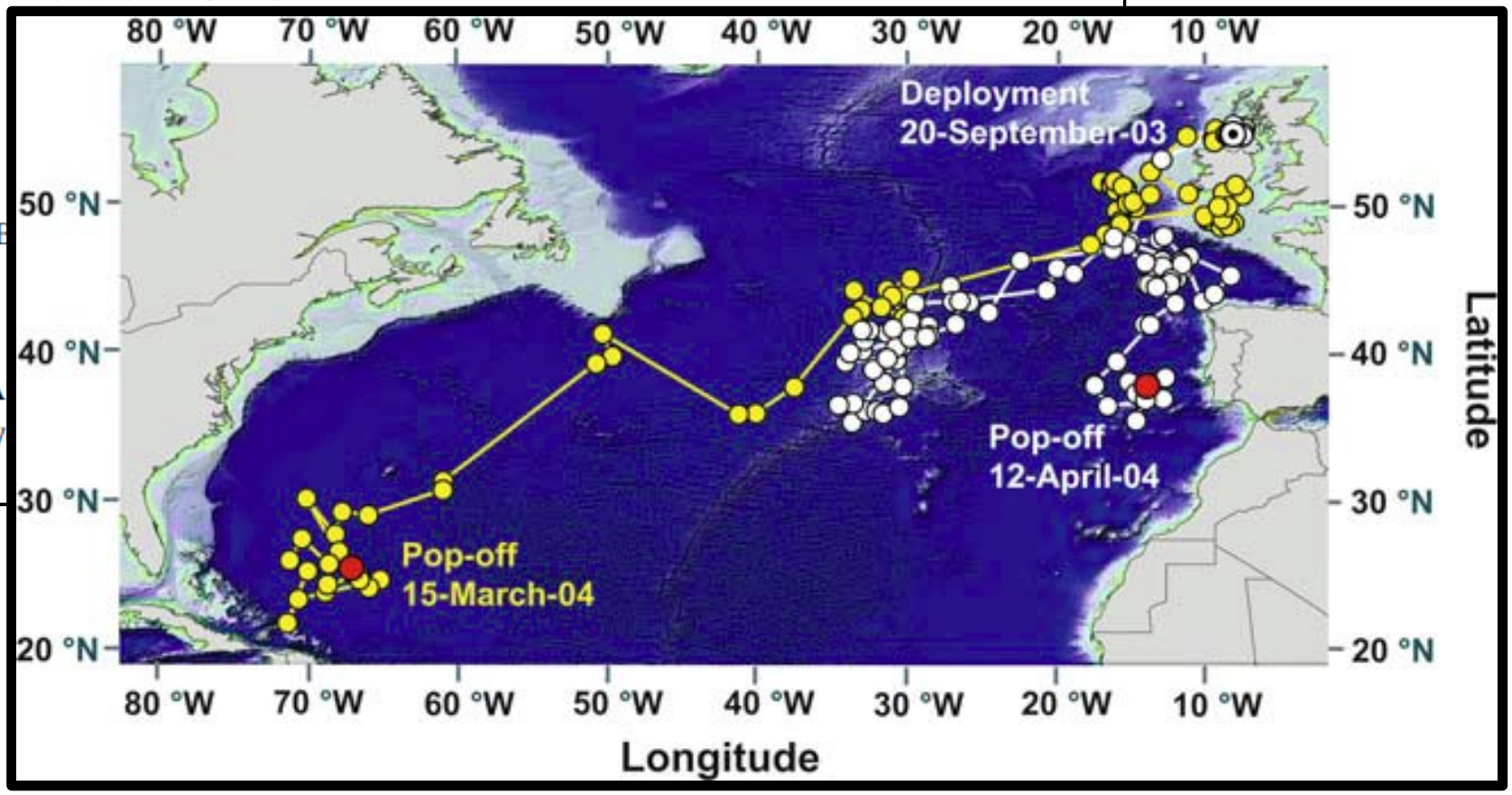
Hydrobiologia (2007) 582:91–97
DOI 10.1007/s10750-006-0552-y

FISH TELEMETRY

Results of satellite tagging of Atlantic bluefin tuna, *Thunnus thynnus*, off the coast of Ireland

Michael J. W. Stokesbury · Ronan Cosgrove · Andre Boustany · Daragh Browne · Steven L. H. Teo · Ronald K. O'Dor · Barbara A. Block

Abstract Pop-up
attached to six A
thynnus) off the w
2003 and 2004.



Noncentral question:

How do tuna spread?

Classic version:

Perform detailed study of diffusion among small group of individuals.

Alternative version:

Make use of an unusual event that makes the typically invisible patterns visible.

Central question:

How do ideas spread?

Classic version:

Perform detailed study of diffusion among small group of individuals.

Alternative version:

Make use of an unusual event that makes the typically invisible patterns visible.

Central question:

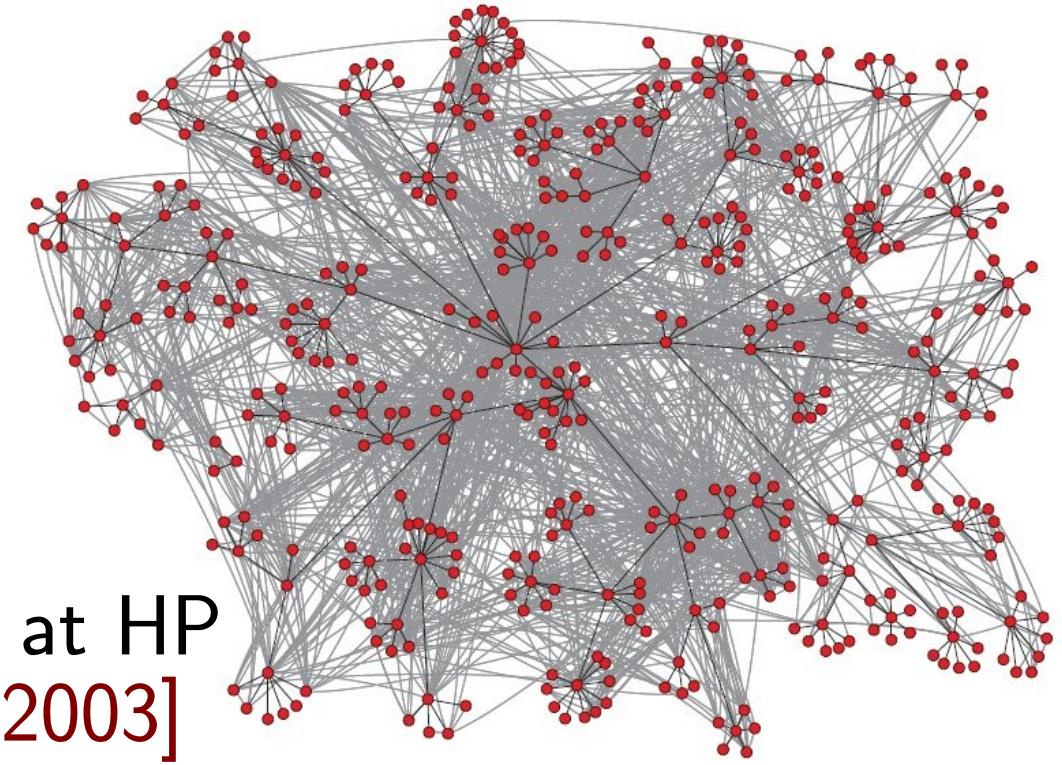
How do ideas spread?

Ideas, disease, innovation, jokes spread continually via the global social network.

But *how?*

What mechanisms?

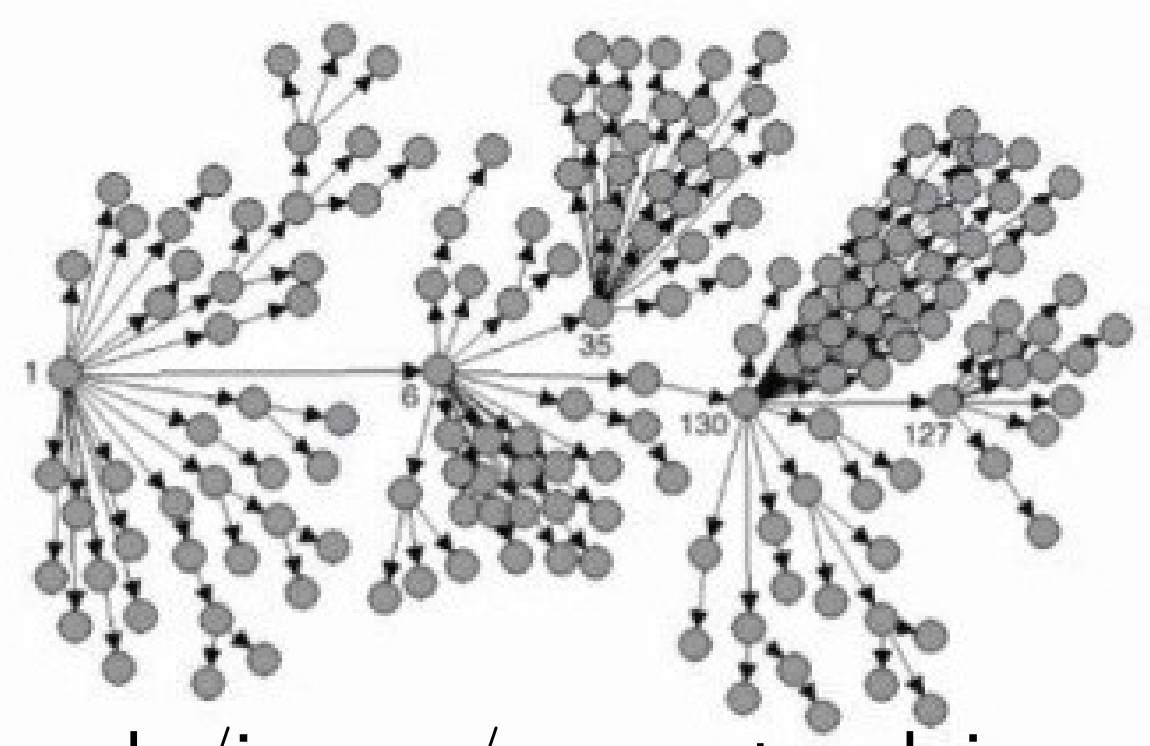
From whence the data?



Email network at HP
[Adamic Adar 2003]

Generally hard to get genuine, large-scale data on a single entity's diffusion.

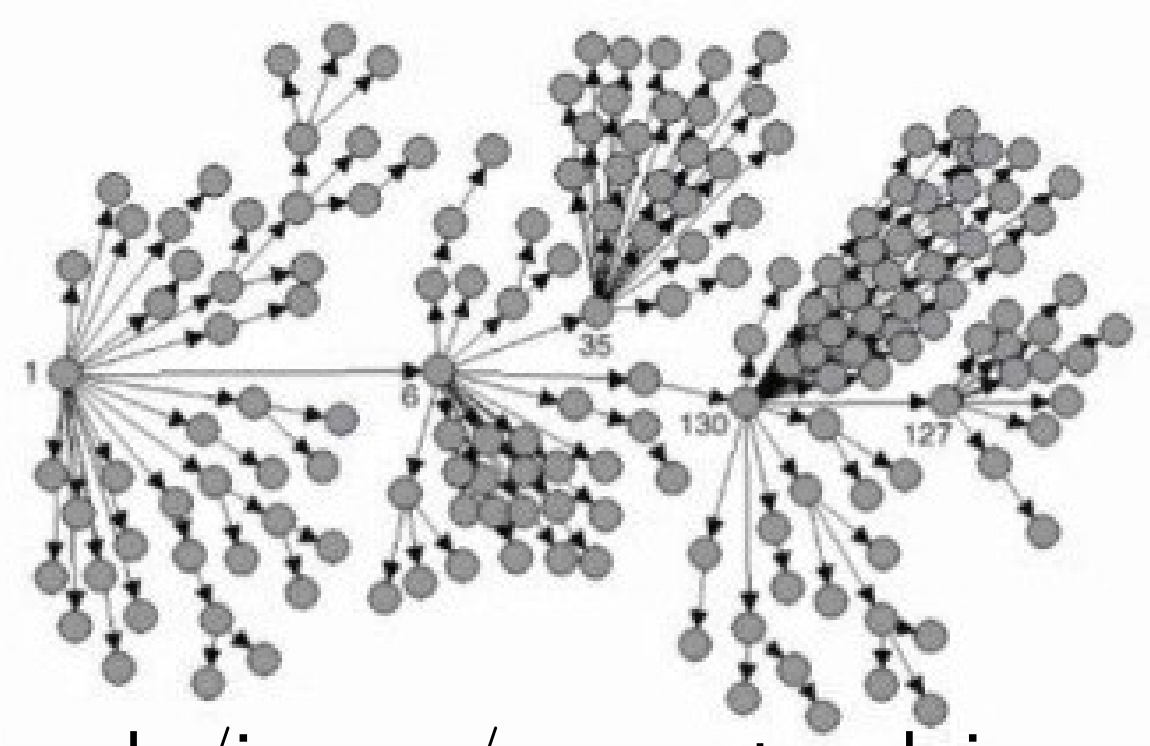
Diffusion of innovation
Contact tracing (epidemiology)
Folklore



http://web.mit.edu/networks/images/sars_network.jpg

Generally hard to get genuine, large-scale data on a single entity's diffusion.
(But it's getting easier ...)

Diffusion of innovation
Contact tracing (epidemiology)
Folklore



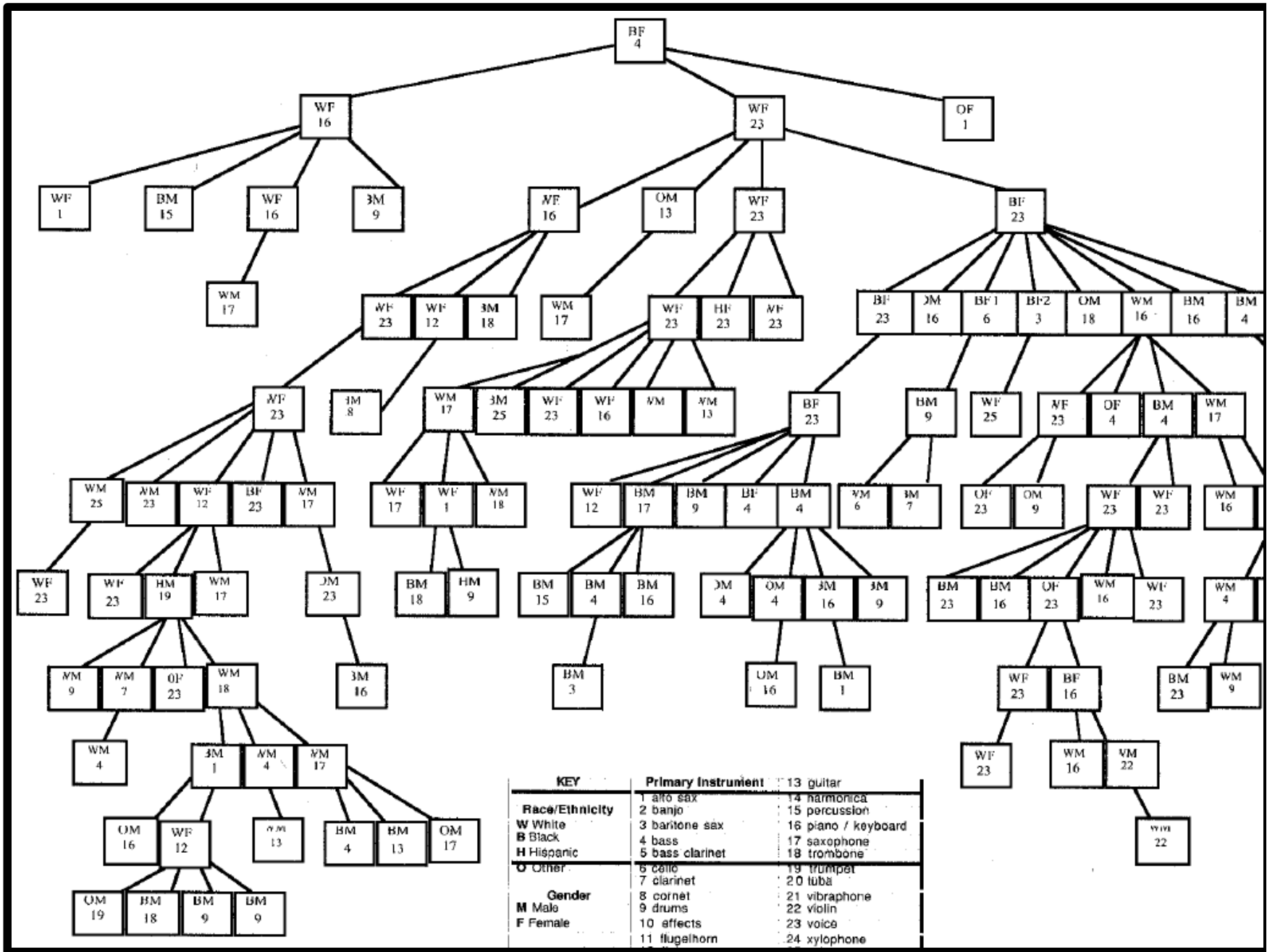
http://web.mit.edu/networks/images/sars_network.jpg

Generally hard to get genuine, large-scale data on a single entity's diffusion.
(But it's getting easier ...)

Diffusion of innovation
Contact tracing (epidemiology)
Folklore
Sampling hidden populations
- [Goodman 1961]
- [Heckathorn 1997]
- [Heckathorn Jeffri 2003]
- ...

Hidden populations: Jazz Musicians in NYC

[Heckathorn Jeffri 2003]



Generally hard to get genuine, large-scale data on a single entity's diffusion.
(But it's getting easier ...)

Diffusion of innovation
Contact tracing (epidemiology)
Folklore
Sampling hidden populations
Inferred spread of topics among blogs/email/Twitter

[Wu Huberman Adamic Tyler 2003]
[Adar Zhang Adamic Lukose 2004] [Adamic Adar 2005]
[Gruhl Guha DLN Tomkins 2004]
[Leskovec McGlohon Faloustos Glance Hurst 2007]
[Kumar Mahdian McGlohon 2010]
[Gomez-Rodriguez Leskovec Krause 2010]
...

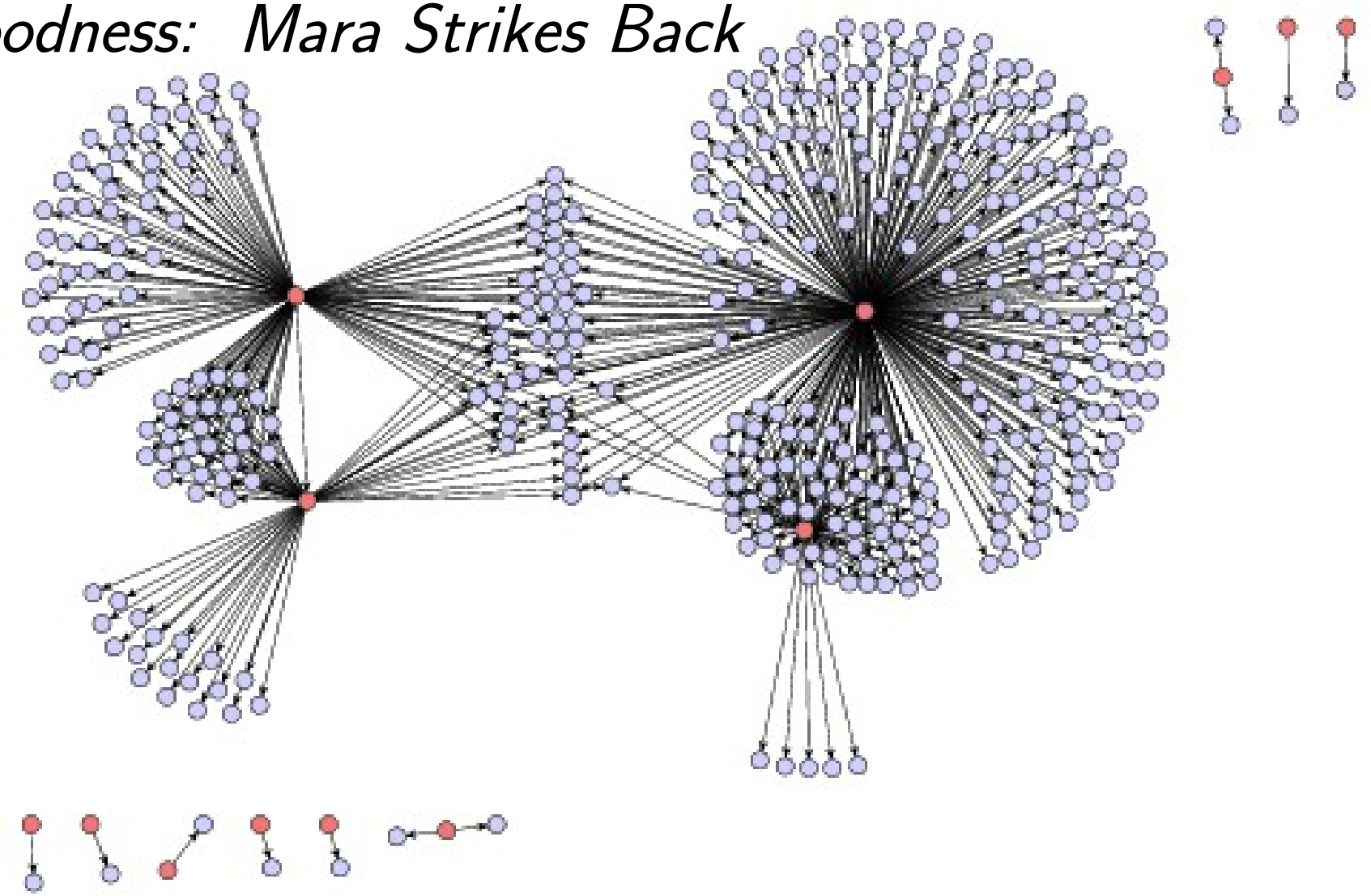
Generally hard to get genuine, large-scale data on a single entity's diffusion.
(But it's getting easier ...)

Diffusion of innovation
Contact tracing (epidemiology)
Folklore
Sampling hidden populations
Inferred spread of topics among blogs/email/Twitter
Word of mouth/viral marketing
- [Goldenberg Libai Muller 2001]
- [Leskovec Adamic Huberman 2006]: product recommendations
- [Iribaren Moro 2009]
- [Sun Rosenn Marlow Lento 2009]: adoption in Facebook feeds
- ...

Viral marketing ("large online retailer")

[Leskovec Adamic Huberman 2006]

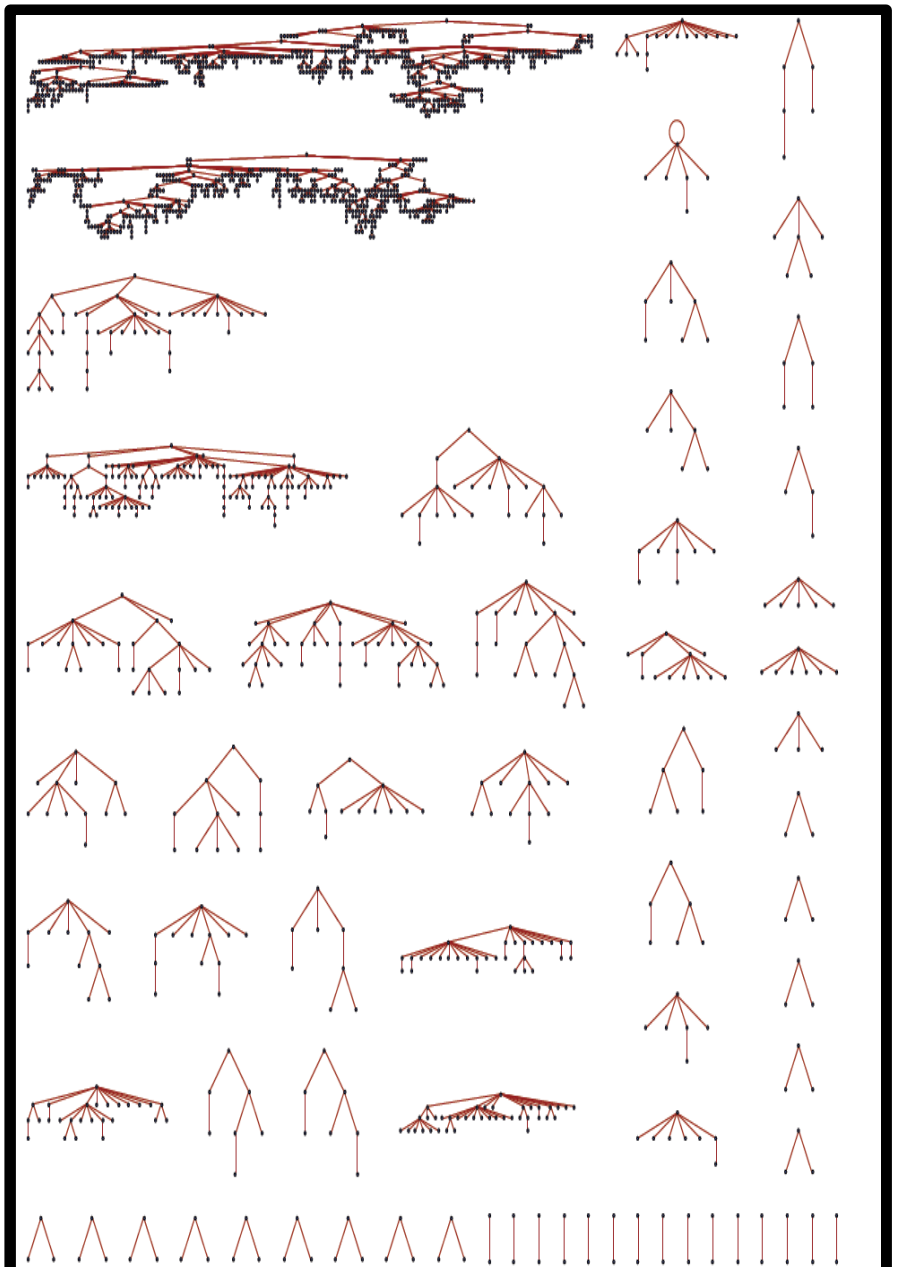
Friends' recommendations for
Oh My Goodness: Mara Strikes Back



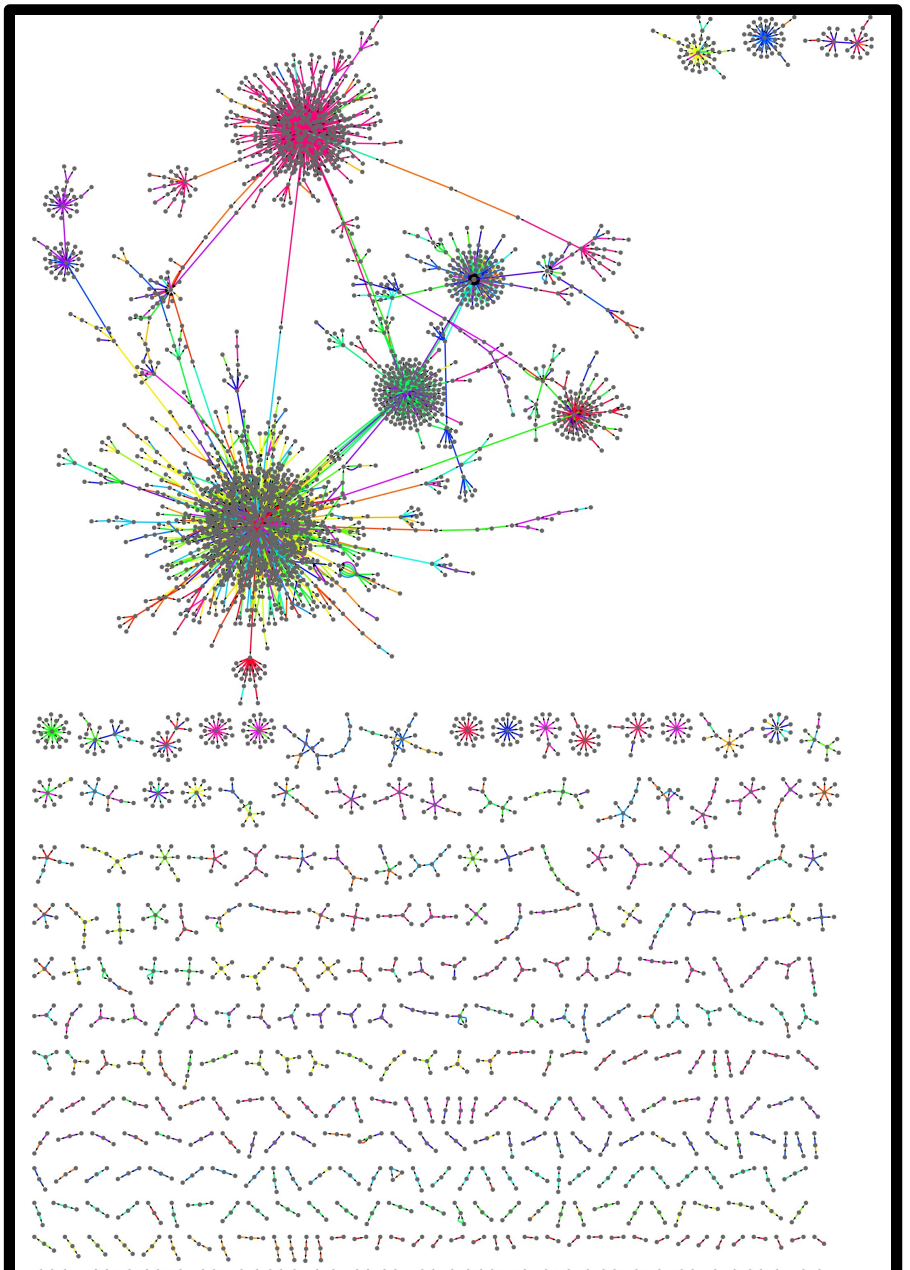
Generally hard to get genuine, large-scale data on a single entity's diffusion.
(But it's getting easier ...)

Diffusion of innovation
Contact tracing (epidemiology)
Folklore
Sampling hidden populations
Inferred spread of topics among blogs/email/Twitter
Word of mouth/viral marketing
Digital traces from online social communities
- [Bakshy Kerrer Adamic 2009]: cloneable assets in Second Life
- [Cha Mislove Gummadi 2009]: Flickr favorites
- [Lerman Ghosh 2010] [Kwak Lee Park Moon 2010]: retweeting
- ...

Gestures in Second Life
[Bakshy Kerrer Adamic 2009]



Retweets in Twitter
[Kwak Lee Park Moon 2010]

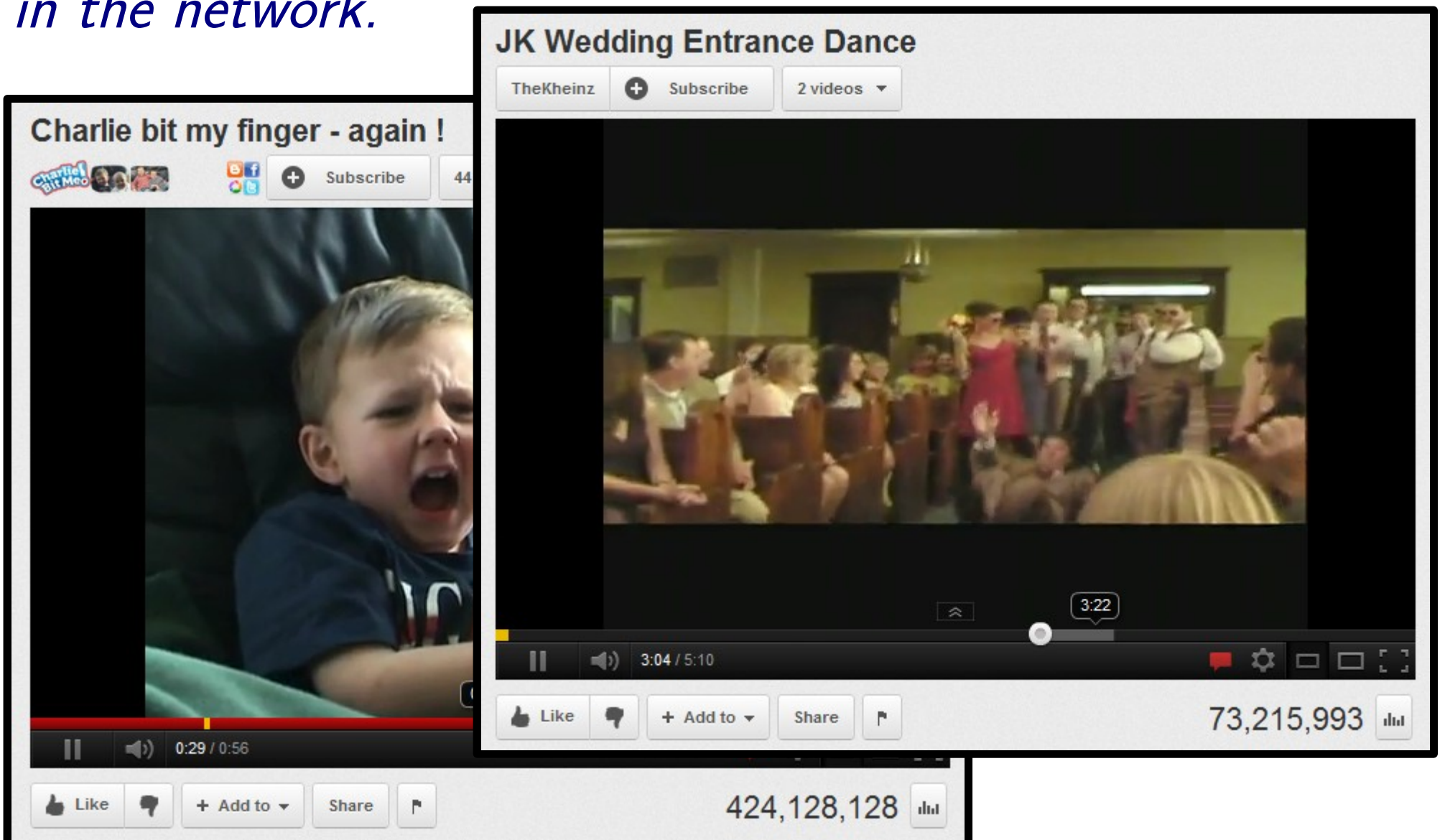


Generally hard to get genuine, large-scale data on a single entity's diffusion.
(But it's getting easier ...)

Diffusion of innovation
Contact tracing (epidemiology)
Folklore
Sampling hidden populations
Inferred spread of topics among blogs/email/Twitter
Word of mouth/viral marketing
Digital traces from online social communities

Intuition: *going viral!*

Some people are susceptible to the meme; it spreads exponentially from "patient 0" through "susceptibles" in the network.



Intuition: *the small-world phenomenon!*

Two people chosen arbitrarily from the social network are connected by a small number of intermediate friends.

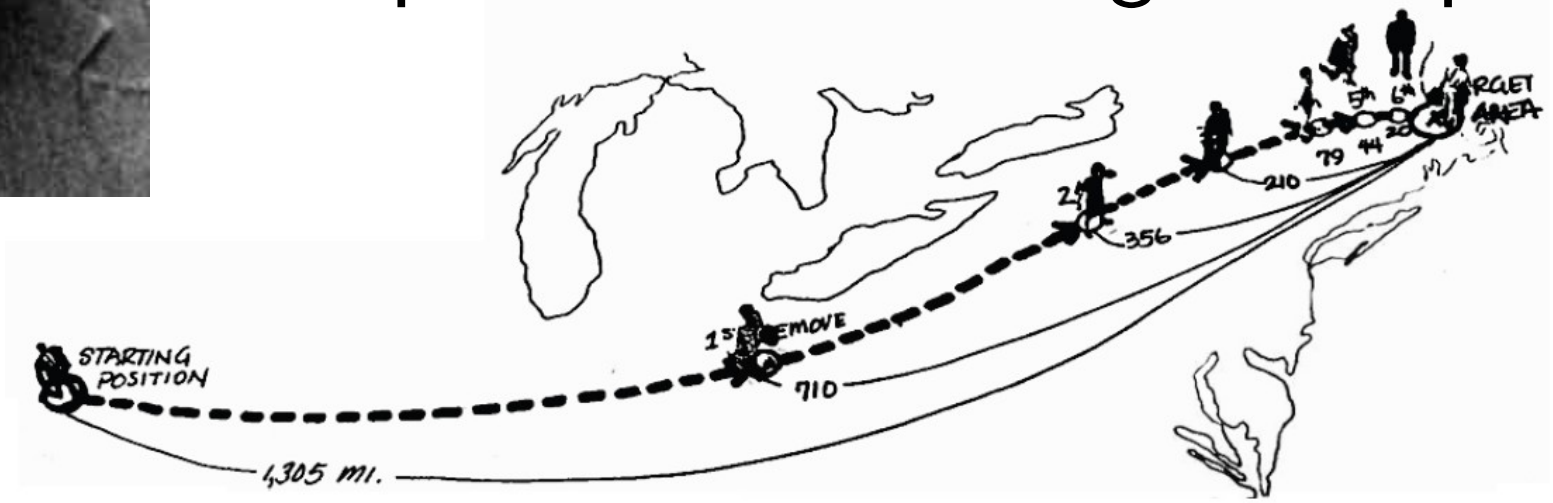
Stanley Milgram [1967]

Participants asked to forward letter to one friend.

source: resident of Omaha, NE

target: stockbroker near Boston

Completed chains averaged 6 hops!

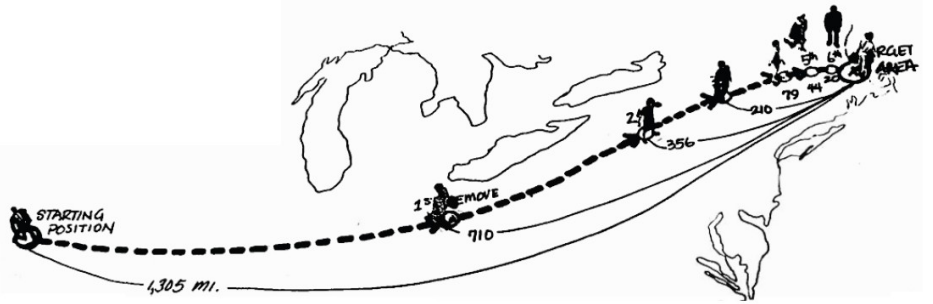


Central question:

How do ideas spread?

Intuition:

- exponential growth ("going viral")
- short chains ("small-world phenomenon")



Central question:

How do ideas spread?

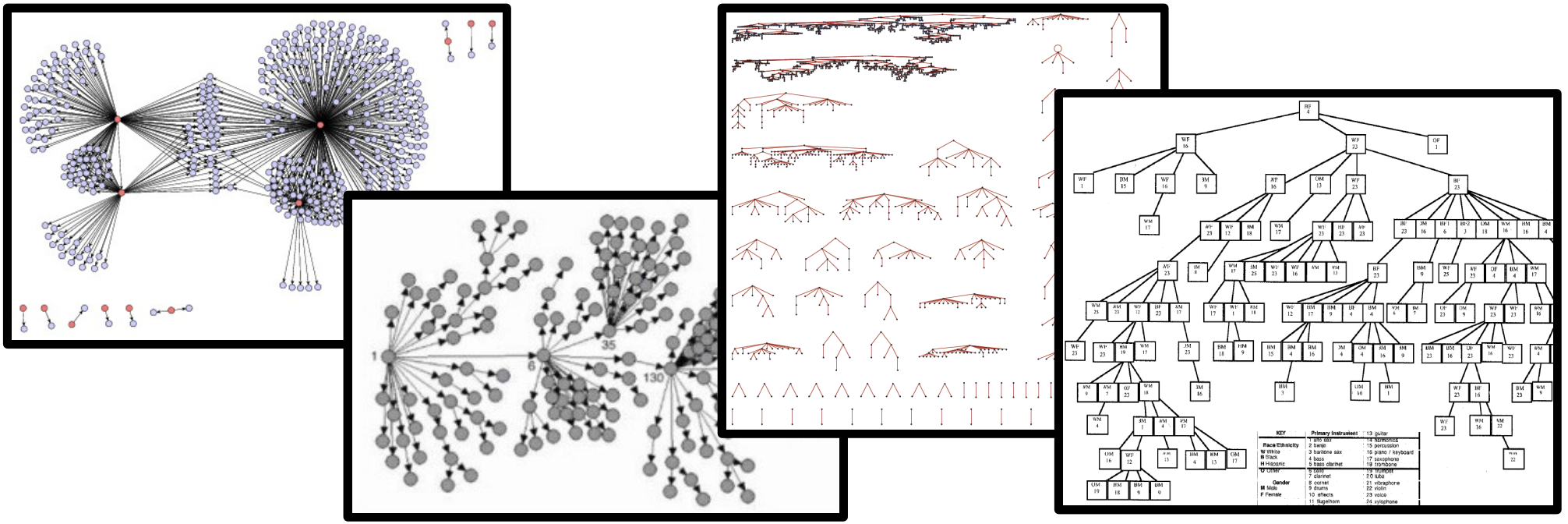
Intuition: exponential growth, short chains

How do we test the intuition?
Where's data on the spread of one idea?

The rest of this talk (a brief summary):

Email-based chain-letter petitions as "tracers" of large-scale propagation through the social network.

[DLN Kleinberg, PNAS 2008]
[Chierichetti Kleinberg DLN, NIPS 2011]



Date: Mon, 17 Mar 2003 16:39:51 -0600
From: XXXX <XXXX@mac.com>
To: usa@un.int, president@whitehouse.gov
Subject: UN Petition

UN Petition for Peace

Non-essential personnel are now evacuating from the US embassies in the middle east. Was is about to start. It takes is 20% of us to cry out for "NO WAR" to induce further diplomacy, but they say our numbers are more like 2%. US Congress has authorized the President of the US to go to war against Iraq. Please consider this an urgent request. UN Petition for Peace, Stand for Peace. Islam is not the Enemy. War is NOT the Answer. Speak against a THIRD WORLD WAR. The UN is gathering signatures in an effort to avoid a tragic world event.

Please COPY (rather than Forward) this e-mail in a new message, sign at the end of the list, and send it to all the people whom you know. If you receive this list with more than 500 names signed, please send a copy of the message to:

usa@un.int and president@whitehouse.gov

Even if you decide not to sign, please consider forwarding the petition on instead of eliminating it

Date: Mon, 17 Mar 2003 16:39:51 -0600
From: XXXX <XXXX@mac.com>
To: usa@un.int, president@whitehouse.gov
Subject: UN Petition

UN Petition for Peace

Non-essential personnel are now evacuating from the US embassies in the middle east. Was is about to start. It takes is 20% of us to cry out [illegible]
say our numbers[illegible]
the President [illegible]
consider this a[illegible]
for Peace. Isla[illegible]
against a THIRD[illegible]
effort to avoid[illegible]

1) Suzanne Dathe, Grenoble, France
2) Laurence COMPARAT, Grenoble, France
3) Philippe MOTTE, Grenoble, France
4) Jok FERRAND, Mont St. Martin, France
5) Emmanuelle PIGNOL, St Martin d'Heres, FRANCE
6) Marie GAUTHIER, Grenoble, FRANCE
7) Laurent VESCALO, Grenoble, FRANCE
[...]

Please COPY (ra[illegible]
sign at the end[illegible]
you know. If you receive this list with more than 500 names signed, please send a copy of the message to:

usa@un.int and president@whitehouse.gov

Even if you decide not to sign, please consider forwarding the petition on instead of eliminating it

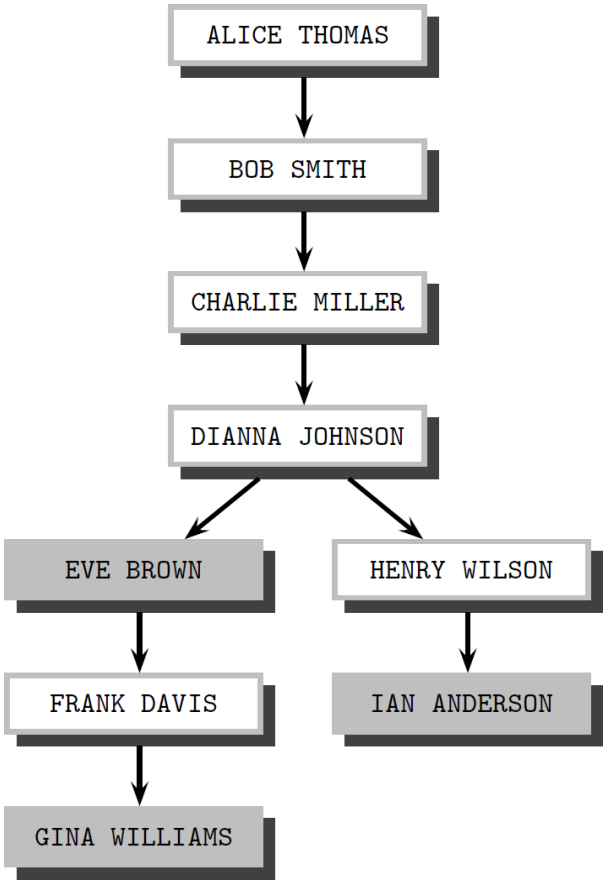
“Before we start”

(80% of the work)

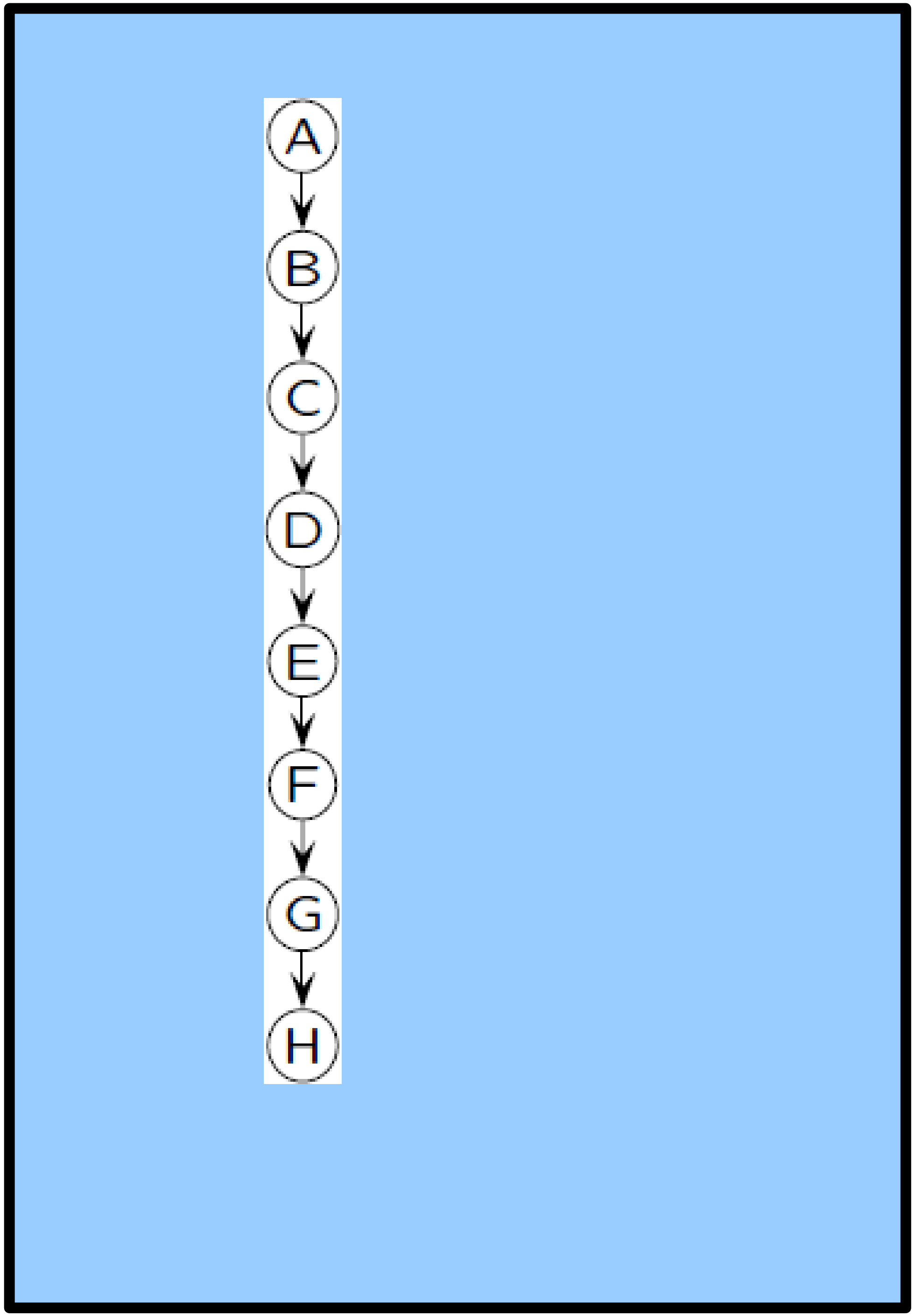
1) Query search engines to find copies of petitions. *(Got 650 distinct copies, ~20K names.)*

1) Query search engines to find copies of petitions. *(Got 650 distinct copies, ~20K names.)*

2) Build propagation tree from copies.
(x,y) edge = x immediately precedes y in some copy

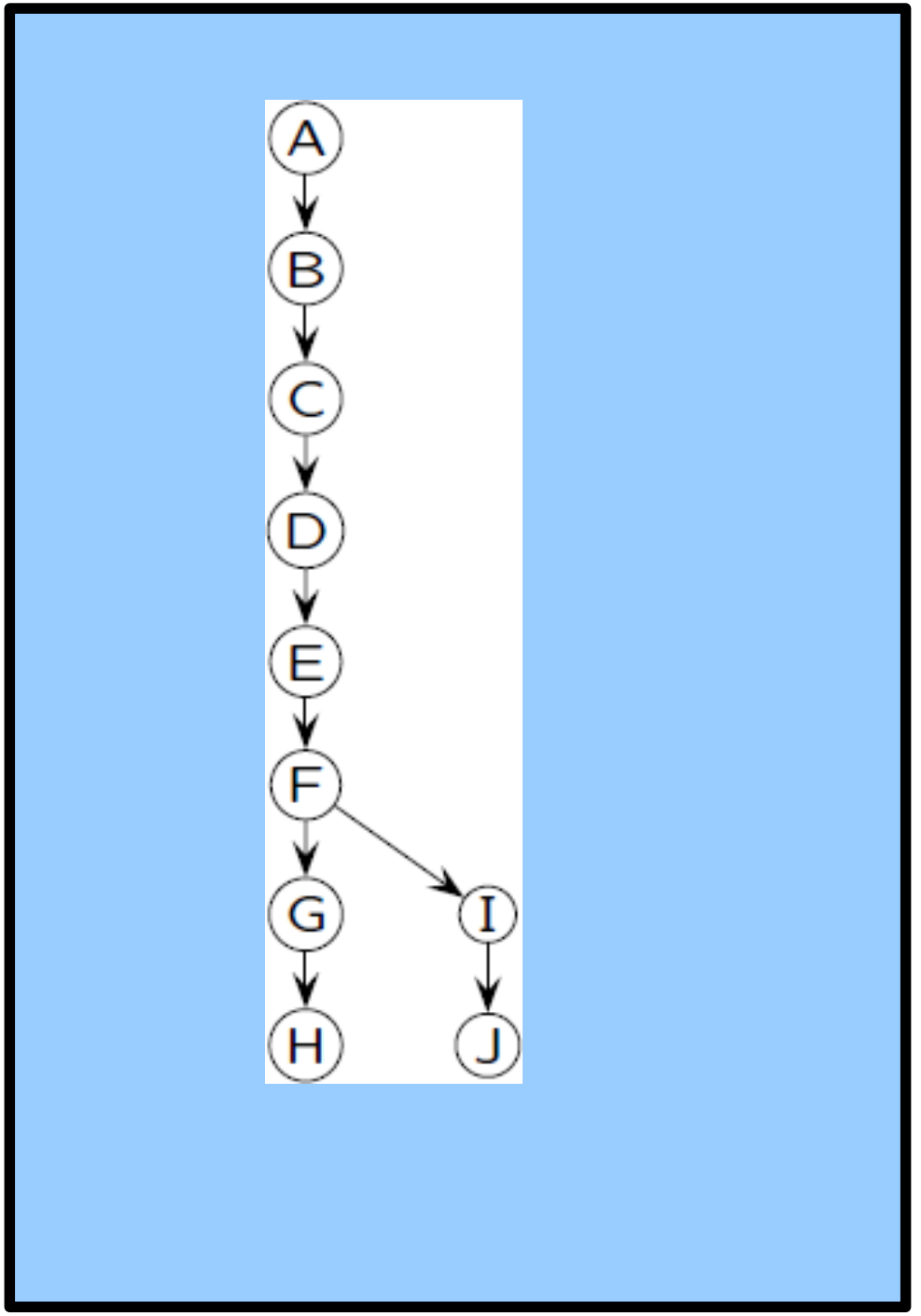


A
B
C
D
E
F
G
H



A
B
C
D
E
F
G
H

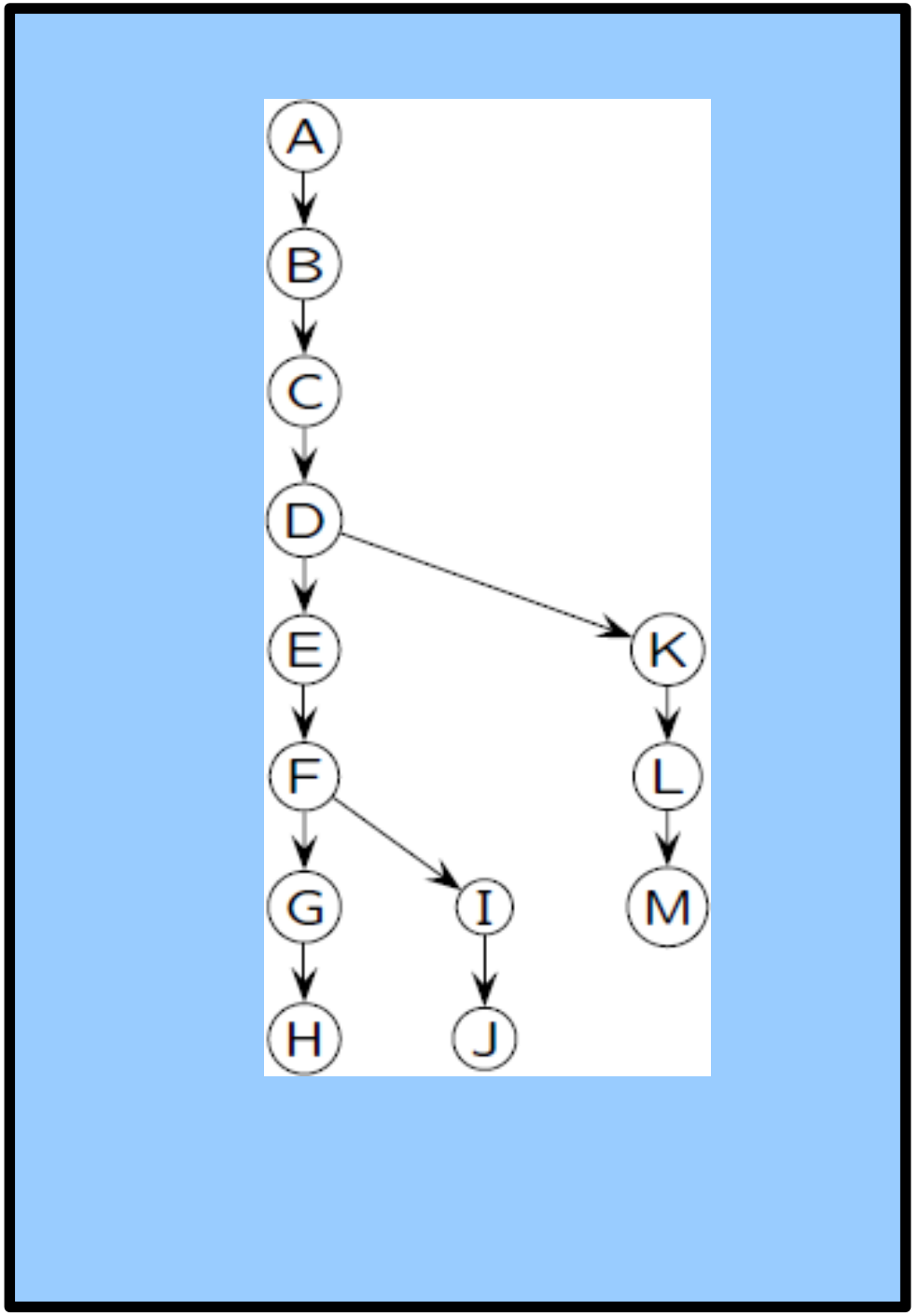
A
B
C
D
E
F
I
J



A B C D E F G H

A B C D E F I J

A B C D K L M



Point mutation: names replaced by political figures

Insertion/deletion: blocks of ~5 names missing/included

Duplication: blocks of ~10 names sometimes repeated

Transposition: two blocks of ~5 names swapped

Hybridization: two lists interleaved in third list(!)

1) Query search engines to find copies of petitions. *(Got 650 distinct copies, ~20K names.)*

2) Build propagation *graph G* from copies.
(x,y) edge = x immediately precedes y in some copy
[typographic variation handled via edit distance threshold]

1) Query search engines to find copies of petitions. *(Got 650 distinct copies, ~20K names.)*

2) Build propagation *graph G* from copies.
(x,y) edge = x immediately precedes y in some copy
[typographic variation handled via edit distance threshold]

3) Delete edges/nodes to form tree T.
[Edmonds 1964 "max weight spanning arborescence"]
["genome rearrangements" handled via pruning]

Central question:

How do ideas spread?

Intuition:

exponential growth ("going viral")

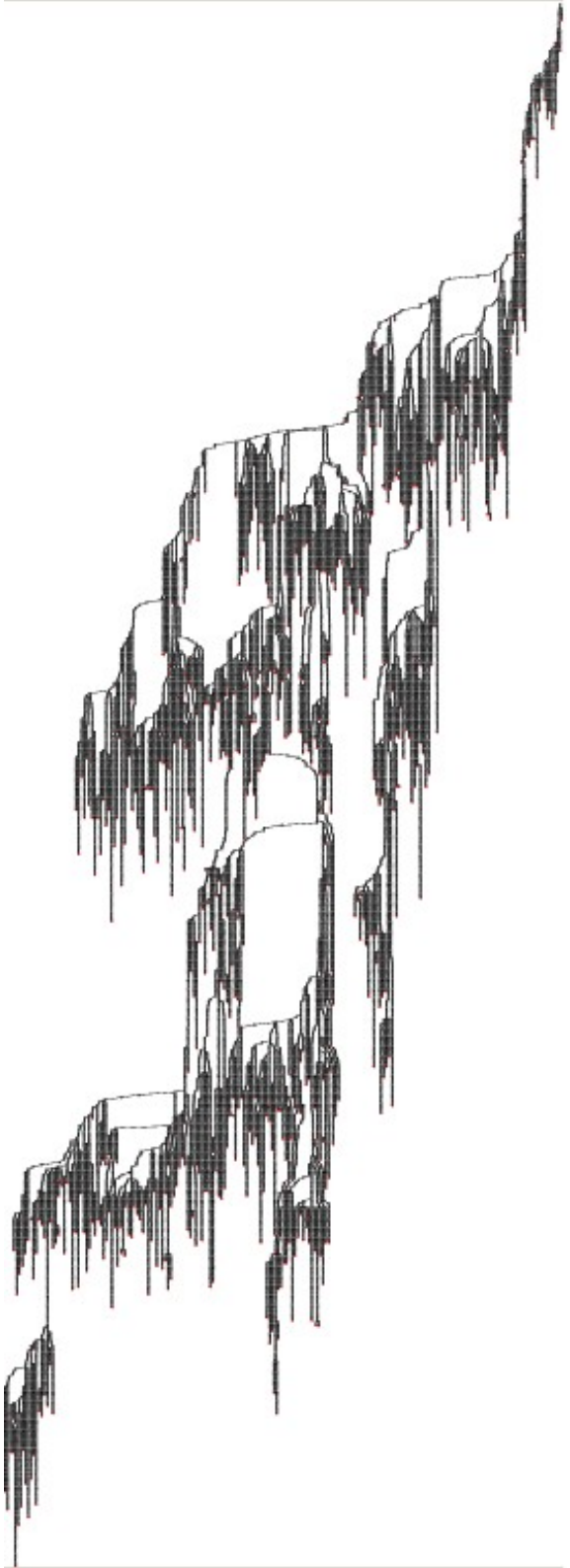
short chains ("small-world phenomenon")

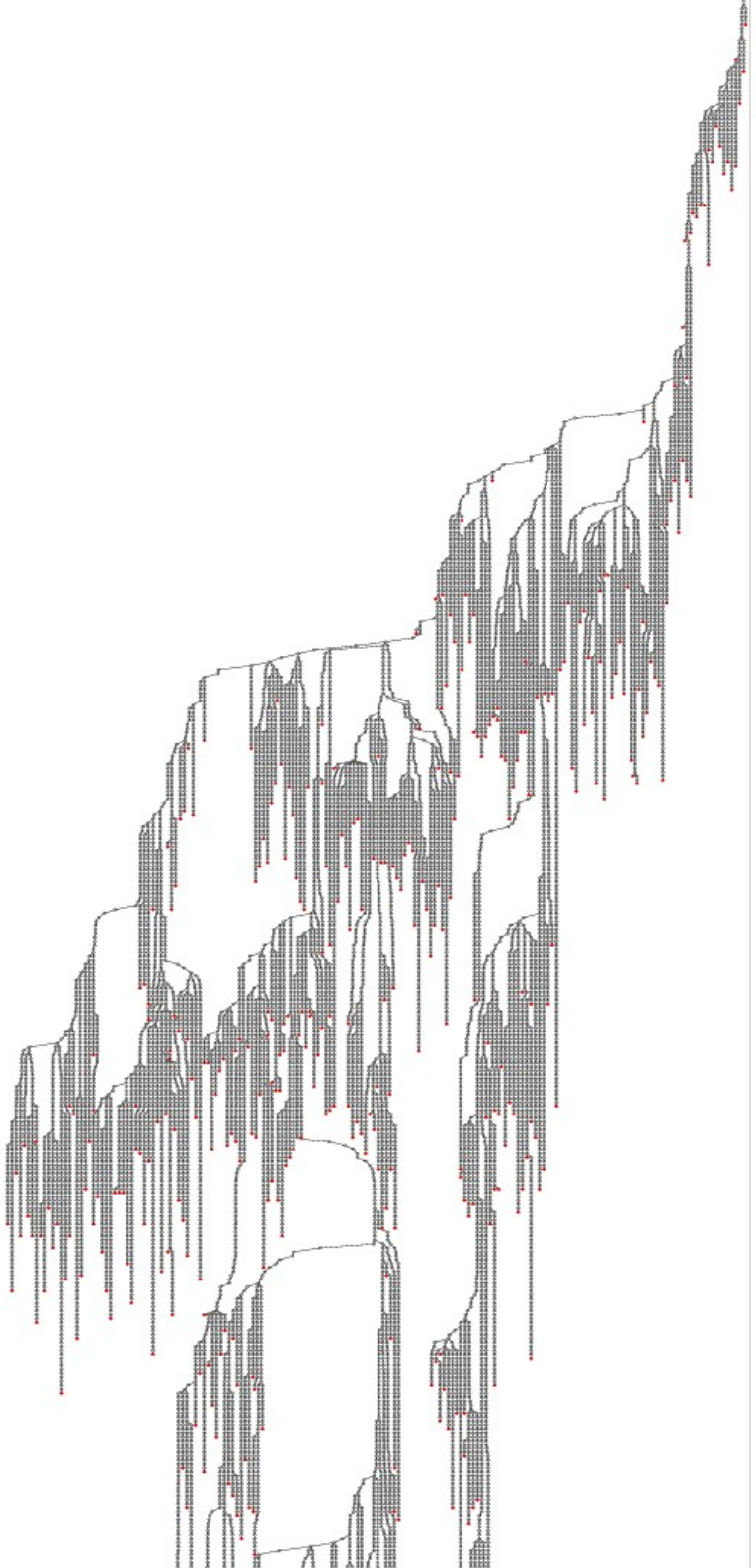
So we'd expect:

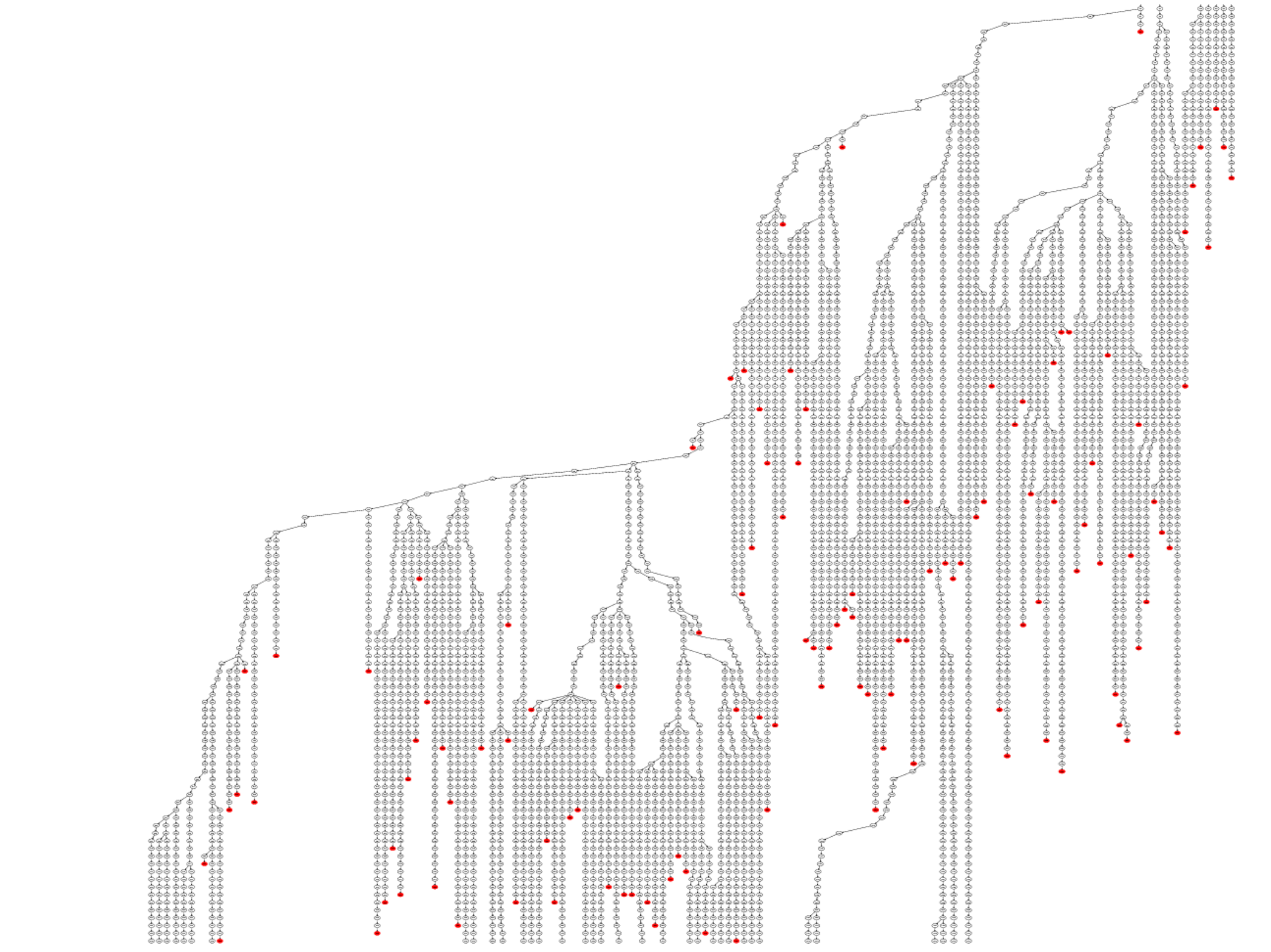
– small depth (small world)

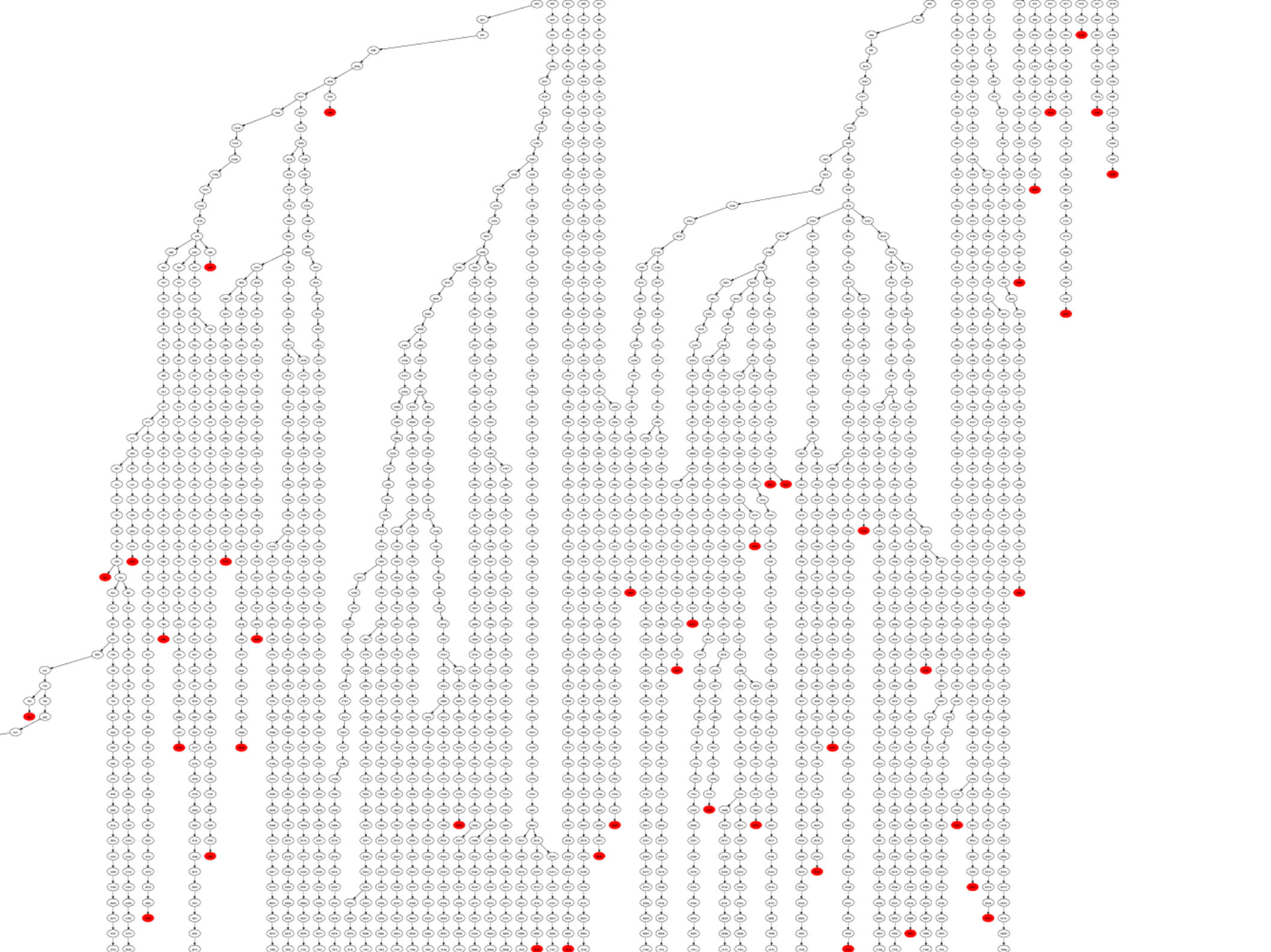
– high branching (10s to 100s of friends)

– *shallow & wide propagation tree*







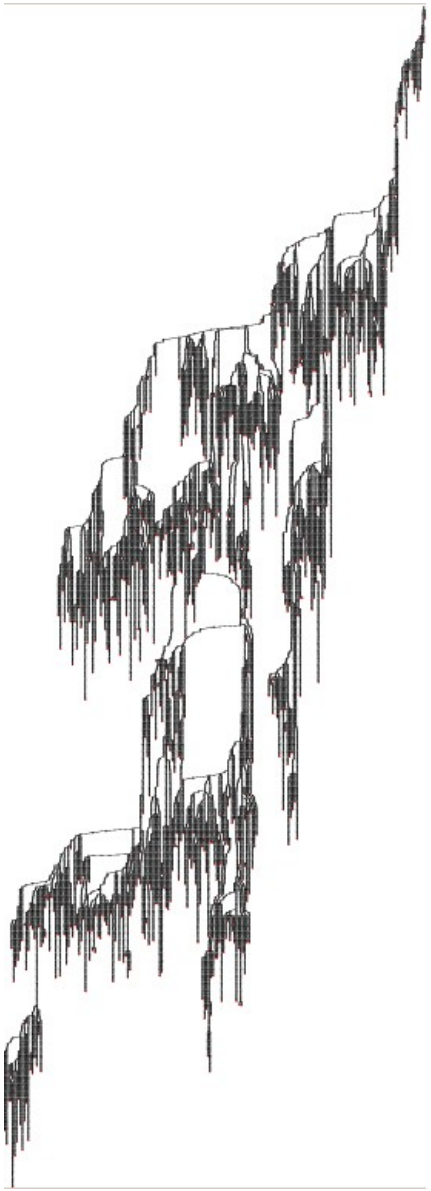


Modeling and Implications

(20% of the work)

Modeling goals:

"good" trees: large median depth, small width, high single-child fraction



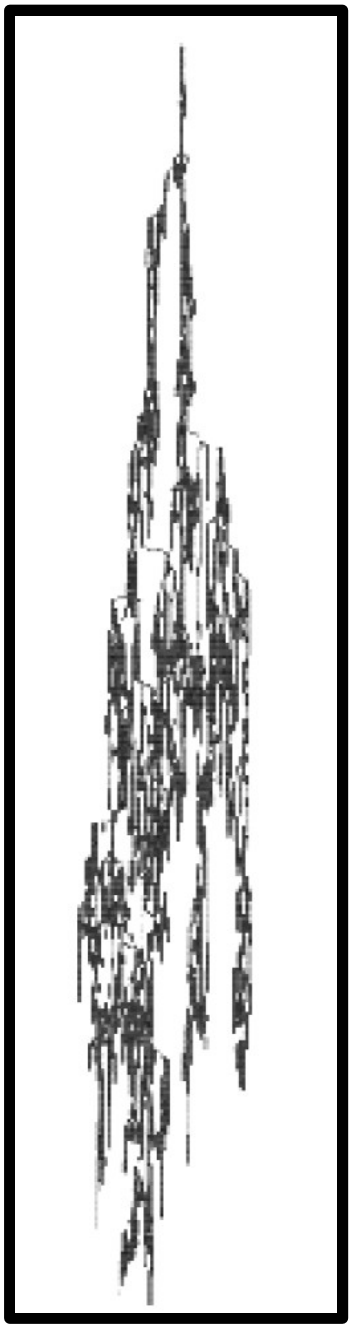
Goal:

simple, plausible generative model that reproduces the observed features.

Model #1: *mechanisms/real networks.*

[DLN Kleinberg 2008]

Test models using real social network data (4.4M LiveJournal nodes).



The epidemic model:

Every non-discarding node forwards to all LJ friends, and posts with some probability p.

Plus variation in timing in responses.

Plus "reply-all" mechanism.

"All models are wrong, but some are useful."

– *George Box*

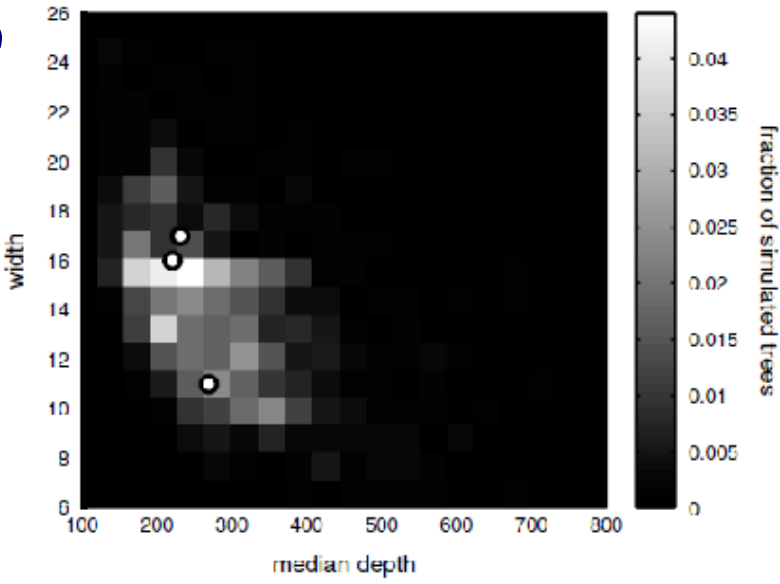
Model #2: *branching processes.*

[Golub Jackson 2010]

D = degree distribution from real Iraq tree.
Define branching process using D.

E[degree] = (n-1)/n, so BP is barely subcritical.

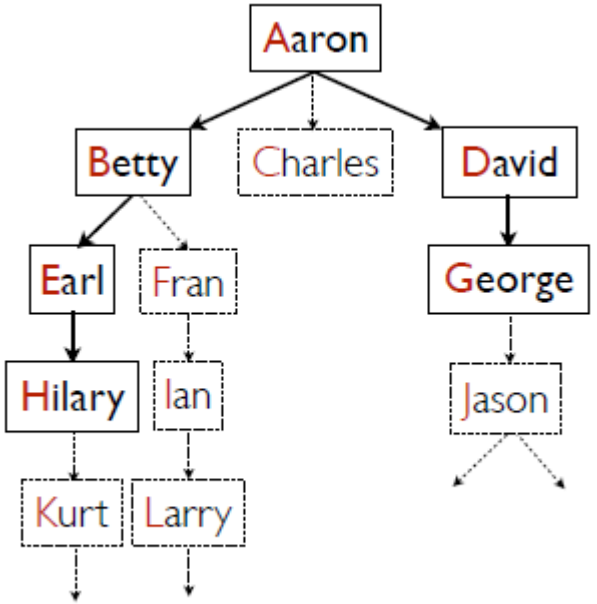
Most generated trees are too small.
But *conditioned on observable tree reaching Iraq size*, simulation shows depth/width of real Iraq tree is typical of trees generated by BP.

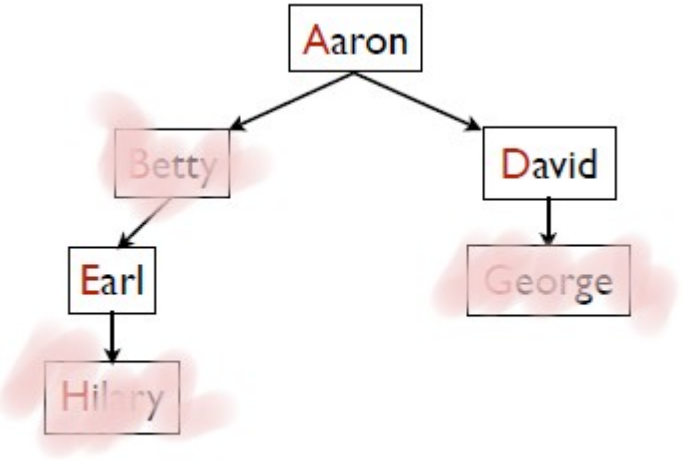


Model #3: *it's all about observation.*

[Chierichetti Kleinberg DLN 2011]

Consider an arbitrary underlying tree T.
Let each node *expose* itself independently with probability p, revealing its path to root.
Let $T[p]$ denote result: observed tree (random).





Model #3: *it's all about observation.*

[Chierichetti Kleinberg DLN 2011]

Theorem:
If *T*'s max degree is bounded and p is small enough, *T[p]*'s single-child fraction is *1–o(1)*.

(In other words, the tree necessarily looks this way because of the way we *observe* it.)

1) finding data on spread of a single piece of information is hard.

2) propagation tree has unexpected structure.

3) we can explain that structure with a model.

1) finding data on spread of a single piece of information is hard.

(chain letters!)

2) propagation tree has unexpected structure.

(narrow, deep, and stringy!)

3) we can explain that structure with a model.

(with a model that is wrong!)

Even the sampling process is oversimplified; posting decisions don't seem independent.

1) finding data on spread of a single piece of information is hard.

(chain letters!)

2) propagation tree has unexpected structure.

(narrow, deep, and stringy!)

3) we can explain that structure with a model.

(with a model that is wrong!)

Even the sampling process is oversimplified; posting decisions don't seem independent.

4) what else can we learn?

(size of underlying propagation? And??)

Information Diffusion

David Liben-Nowell

dlibenno@carleton.edu

Carleton College, Department of Computer Science

Thank you!