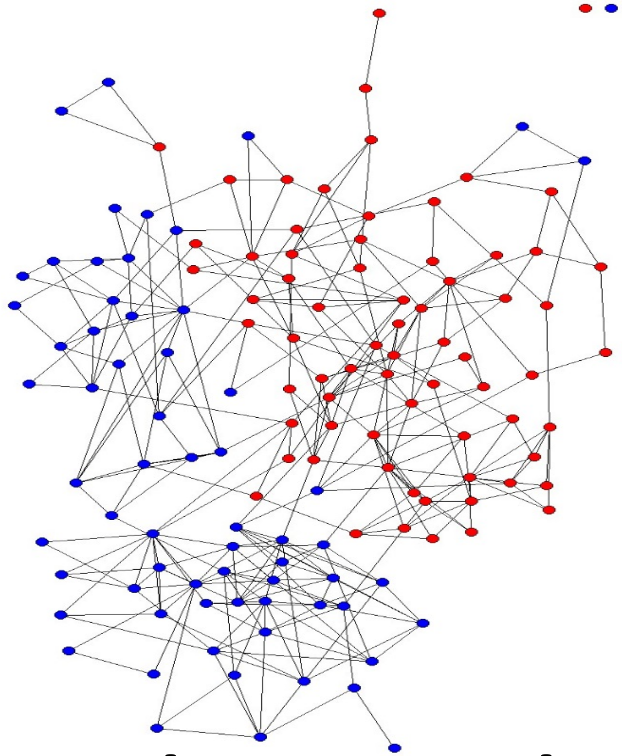
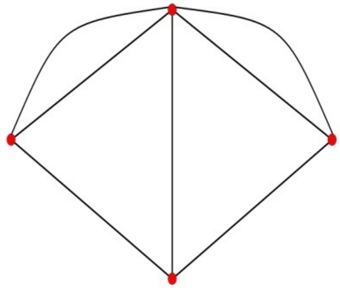
Learning Through the Grapevine:

The Impact of Noise and the Breadth and Depth of Social Networks



Matthew O. Jackson, Suraj Malladi, David McAdams

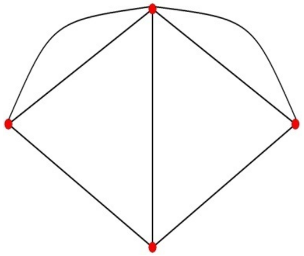
Fighting Disinformation



- *"The safety of our democracy is more important than shareholder dividends and CEO salaries, and we need tech companies to behave accordingly. That's why I'm calling on them to take real steps right now to fight disinformation."* -- Elizabeth Warren

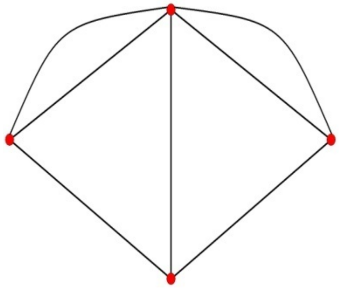
- *"I don't think that Facebook or internet platforms in general should be arbiters of truth"* —Mark Zuckerberg

Avoiding Censoring



- Platforms, crowd-sources, gvts – all have biases
- Huge volume – how to arbitrate `truth’
- Algorithms have hard time with satire, humor, and who is to decide what is ``truth’’?
- How can we design platforms to ensure truth, without policing content?

Preview

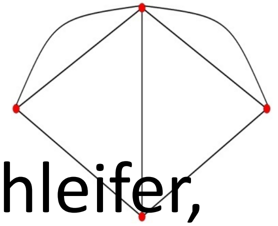


- Learning with noise requires hearing from exponentially many sources

- With tiny uncertainty about `mutation’ rates, limit learning is completely precluded

- Trimming the depth and/or breadth of a network can improve its learning properties

Literature



- Social learning: Banerjee (1992) Bikhchandani, Hirshleifer, Welch (1992), Ellison,Fudenberg (1995), Smith, Sørensen (2000), DeMarzo, Vayanos, Zwiebel (2003), Acemoglu, Ozdaglar, ParendehGheib (2010), Hagenbach, Koessler (2010), Golub and Jackson (2010,2012), Acemoglu, Dahleh, Lobel, Ozdaglar (2011), Mossel, Sly, Tamuz (2015), Bloch, Demange, Kranton (2016), Adamic, Lento, Adar, Ng (2016), Mossel, Mueller-Frank, Sly, Tamuz (2017), Cheng, Hann-Caruthers, Tamuz (2018), Bohren Hauser (2018) ...
- Branching processes: Liben-Nowell and Kleinberg (2008), Golub and Jackson (2010b), Sadler (2017)
- Information theory: Shannon (1947)...
- Birth-Death Models, Evolution Models, `Gossip Algorithms'

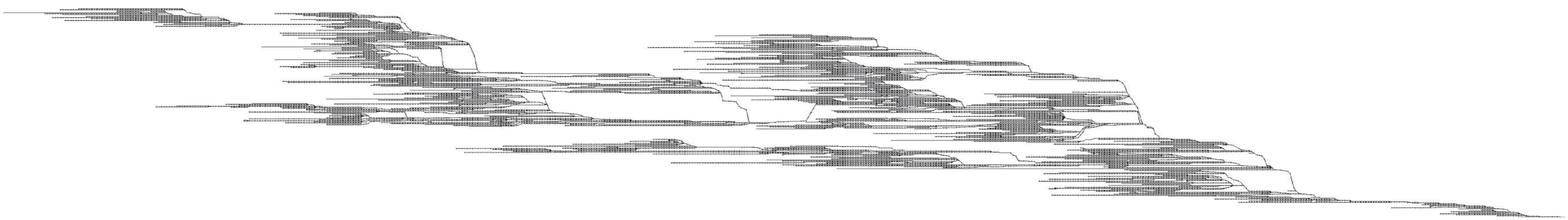
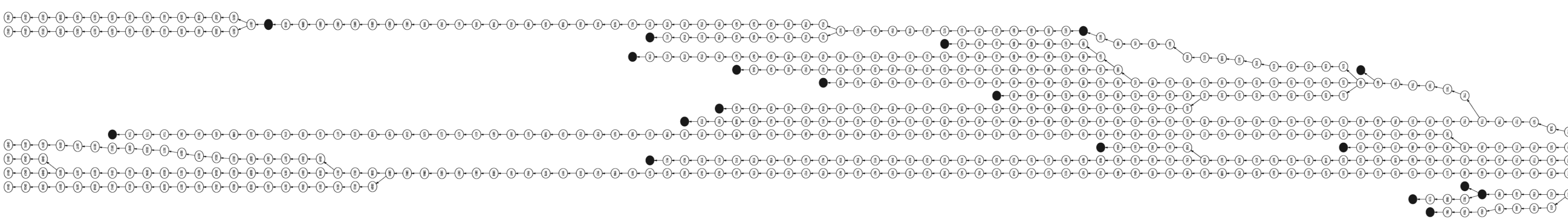
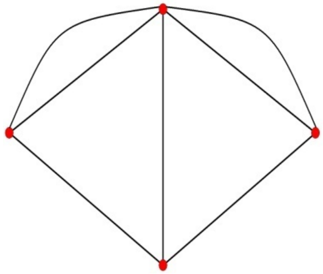


Fig. 2. Tree derived from a large-scale chain-letter petition protesting the start of the war in Iraq, produced as described in Fig. 1. This tree has 18,119 nodes, of which 17,079 (94.26%) have exactly 1 child. The median node depth is 288 and the width of the tree is 82.



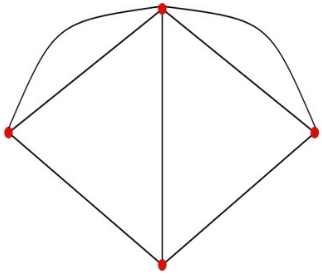
Liben-Nowell and Kleinberg (2008)
Long chains of relayed messages

Example of Mutations…



- Simmons, Adamic, Adar (2011) – tweets:
- “Street style shooting in Oxford Circus for ASOS and Diet Coke. Let me know if you're around!”

Example of Mutations...



- Simmons, Adamic, Adar (2011) – tweets:
- “Street style shooting in Oxford Circus for ASOS and Diet Coke. Let me know if you're around!”

became

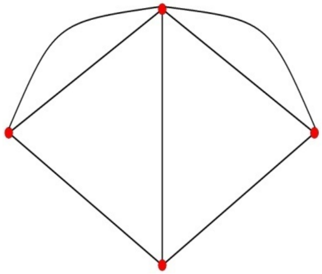
- “Shooting in progress in Oxford Circus? What?"

become

- “Shooting in progress in Oxford Circus, stay safe people.”

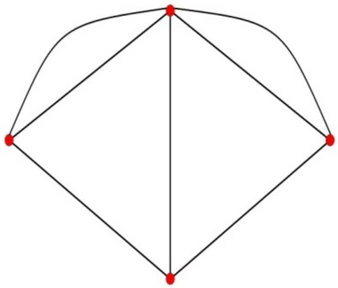
within 3 minutes...

Example of Mutations...



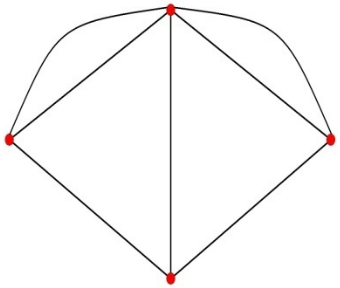
- Adamic, Lento, Adar, Ng (2016) – Facebook memes
- Example: “No one should die because they cannot afford health care and no one should go broke because they get sick. If you agree please post this as your status for the rest of the day.”
- Posted more than 470K times. Mutation rate 11 percent. More than 100K variants...
- Other memes: 121 of 123 largest had more than 100K variants

Outline



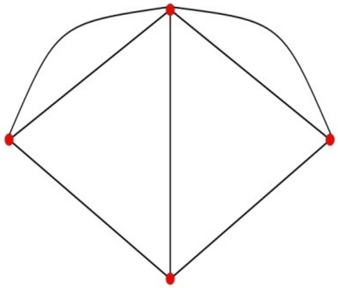
- Model
- Limit learning in long chains
- Optimal regulation/limits on the graph

Model



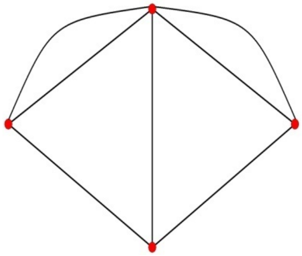
- State ω: either 1 or 0
- Probability: θ that $\omega = 1$
- Initial sender(s) see signal of state
- Receiver hears signal(s) via word of mouth

Mutations



- Prob μ_{01} , μ_{10} < ½ of mutating in transmission
- iid probabilities of mutating
- Could be random, but might also be deliberate
- Regardless of incentives, what matters: μ_{01} , μ_{10}

Transmission Failure



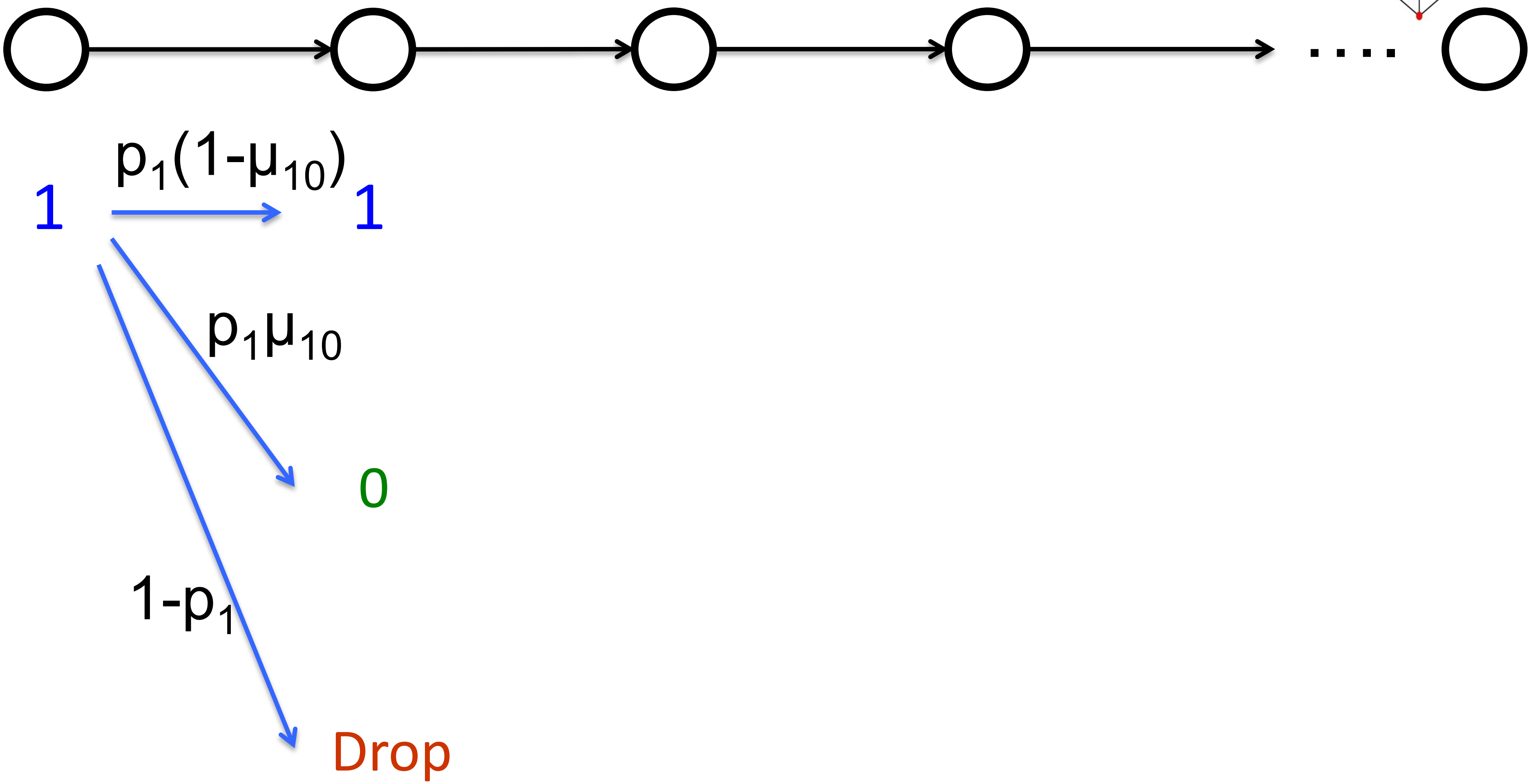
- probability p_1 of passing 1, dropped w prob $(1-p_1)$
- probability p_0 of passing 0, dropped w prob $(1-p_0)$
- E.g.,
  - report `news' – significant results with higher prob.
  - Tell funnier jokes,
  - Pass salacious rumors…
- *Today* $p_1 = p_0 = p$

Passing along a chain

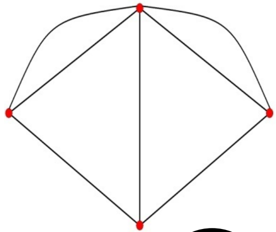
Source/
original sender

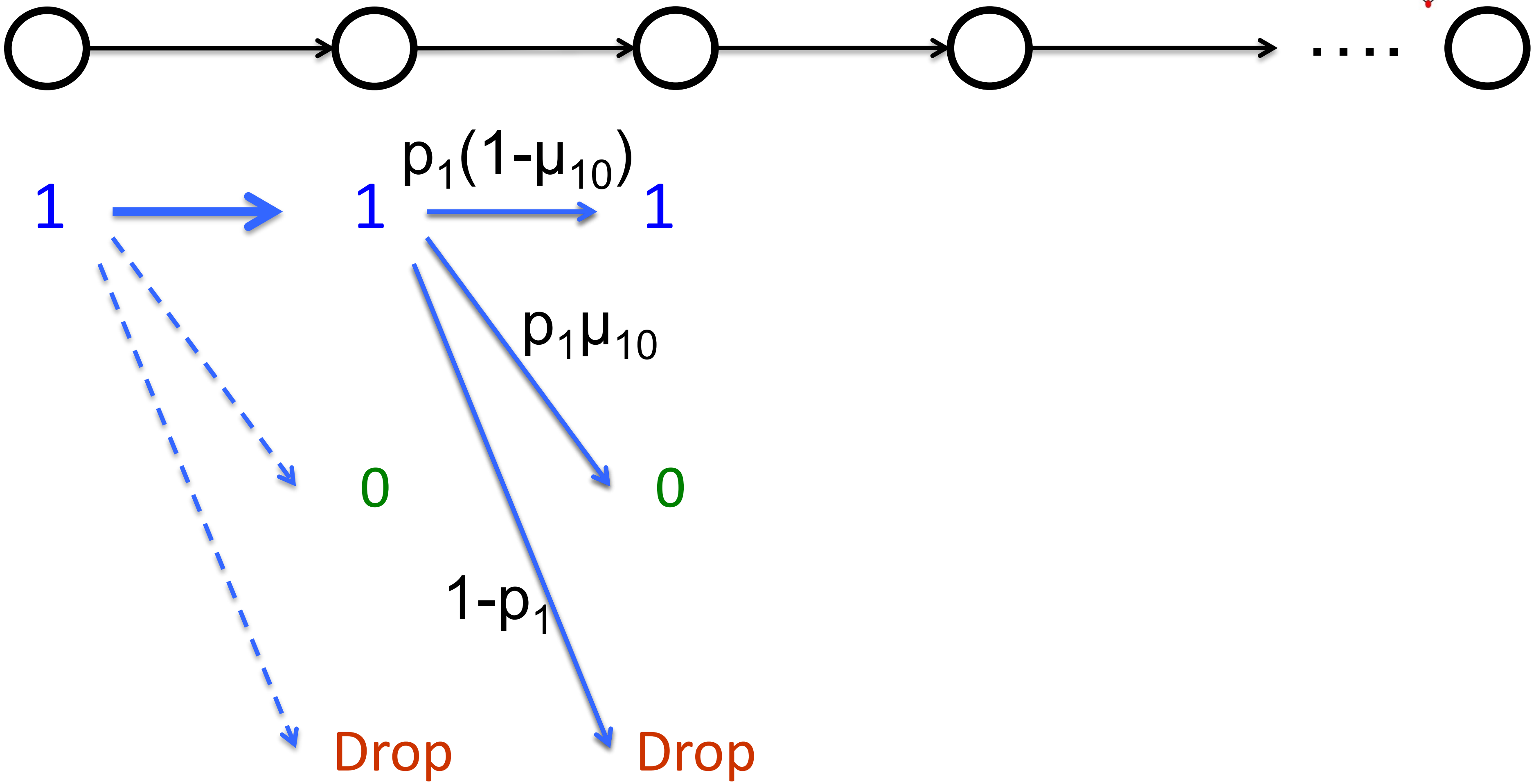
Receiver/
learner

Passing

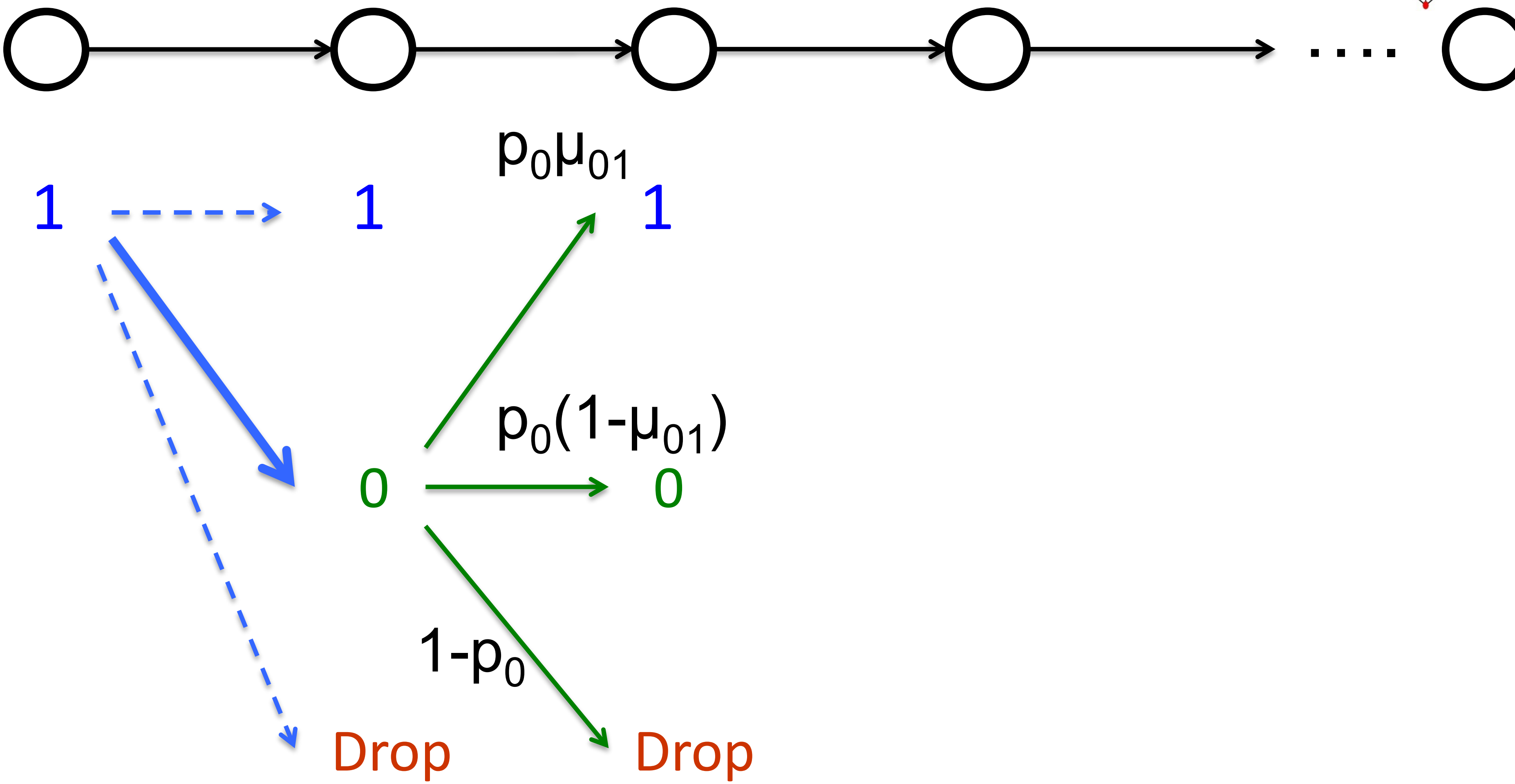


Passing

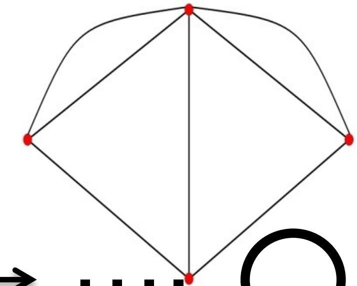


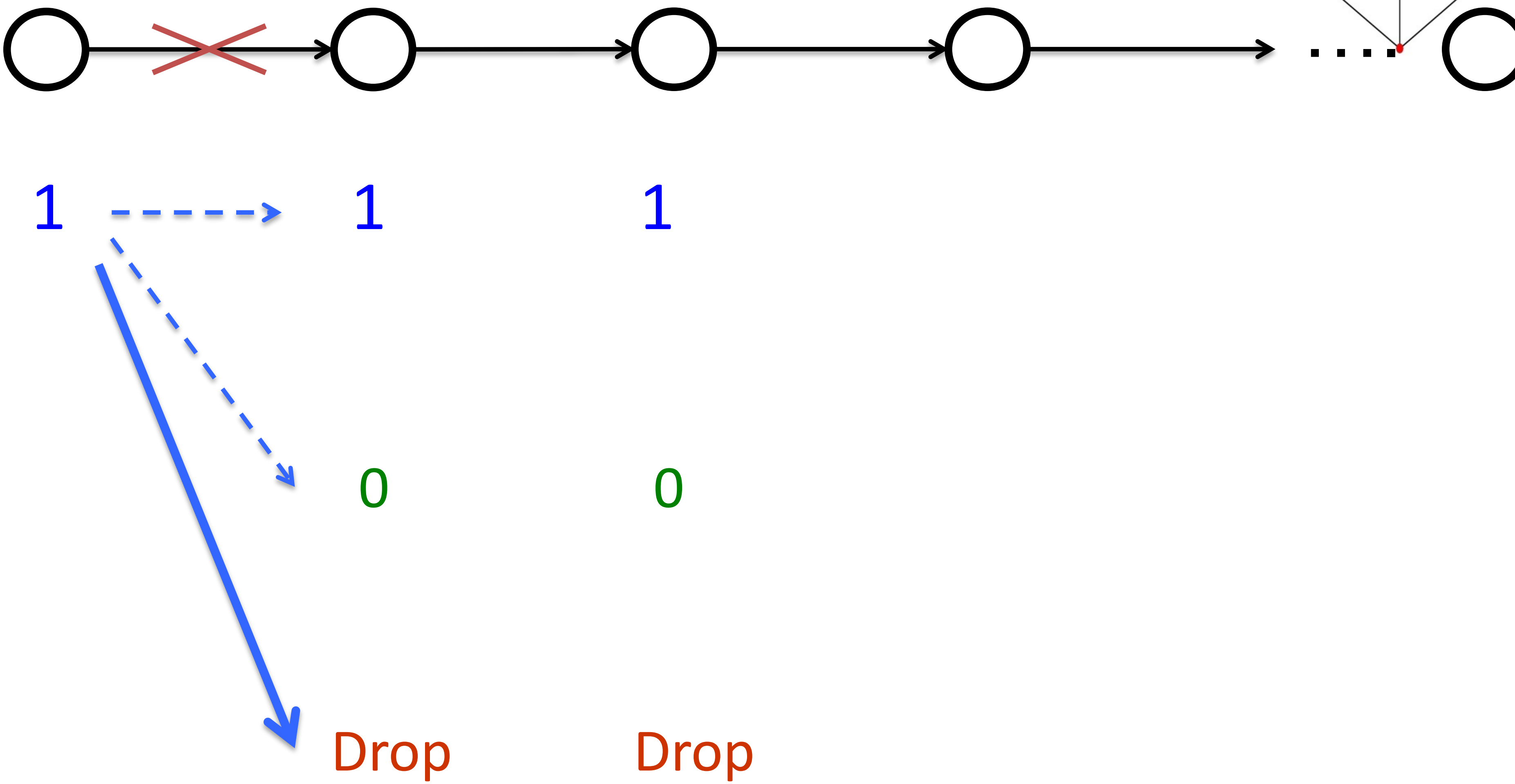


Passing

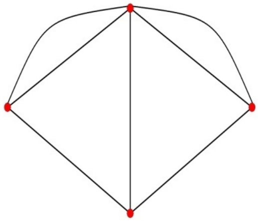


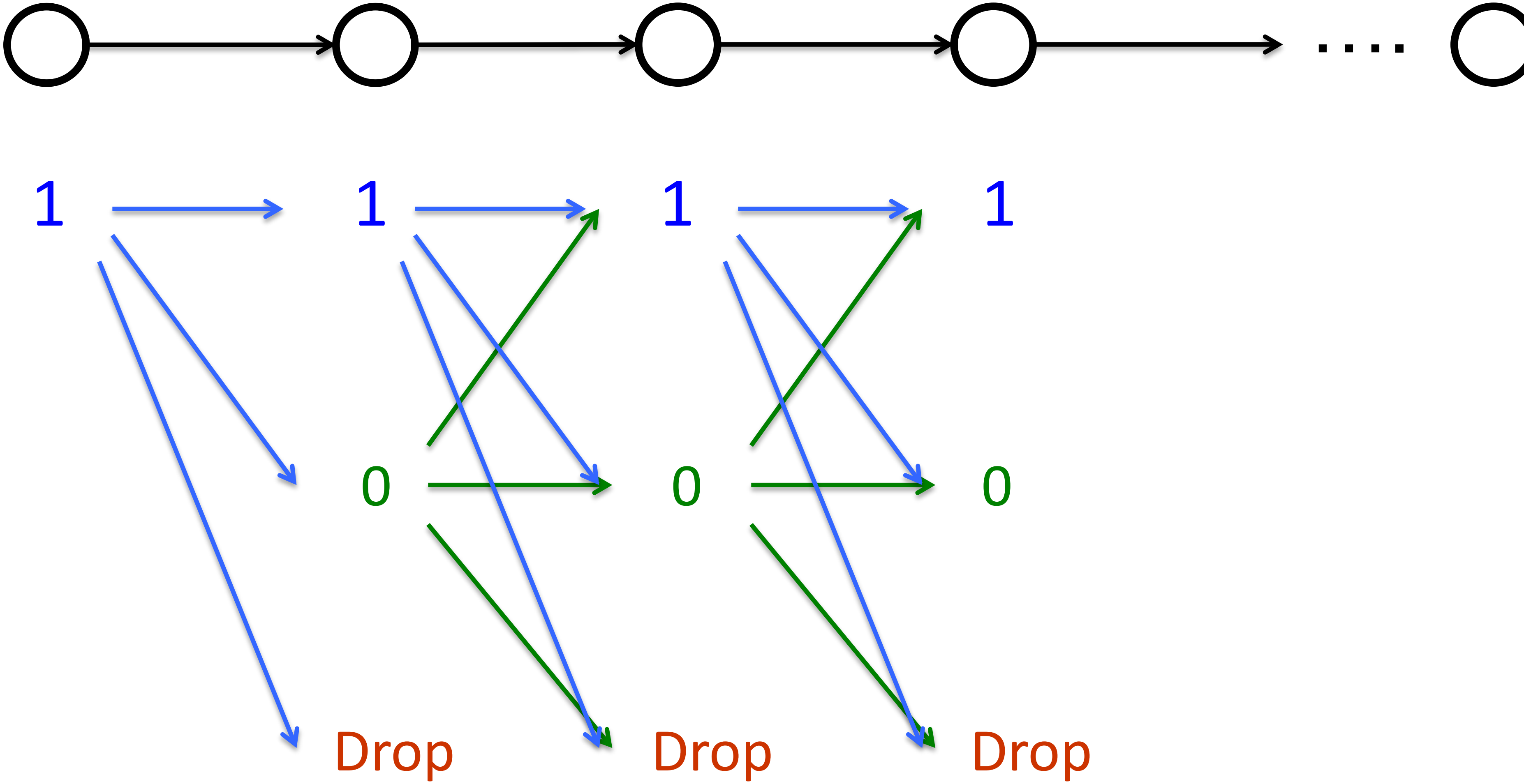
Passing



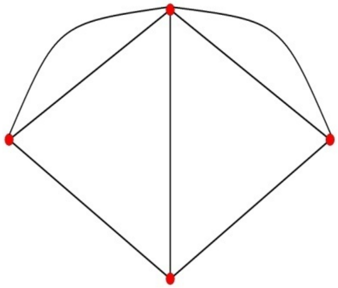


Passing





Chain of passing

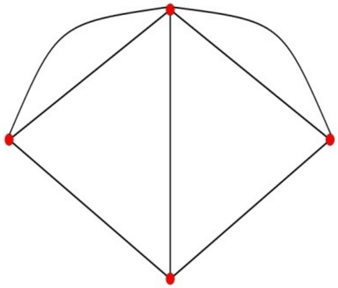


- t in {1, 2, 3, ...}

- Markov chain with states 1, 0, ϕ:

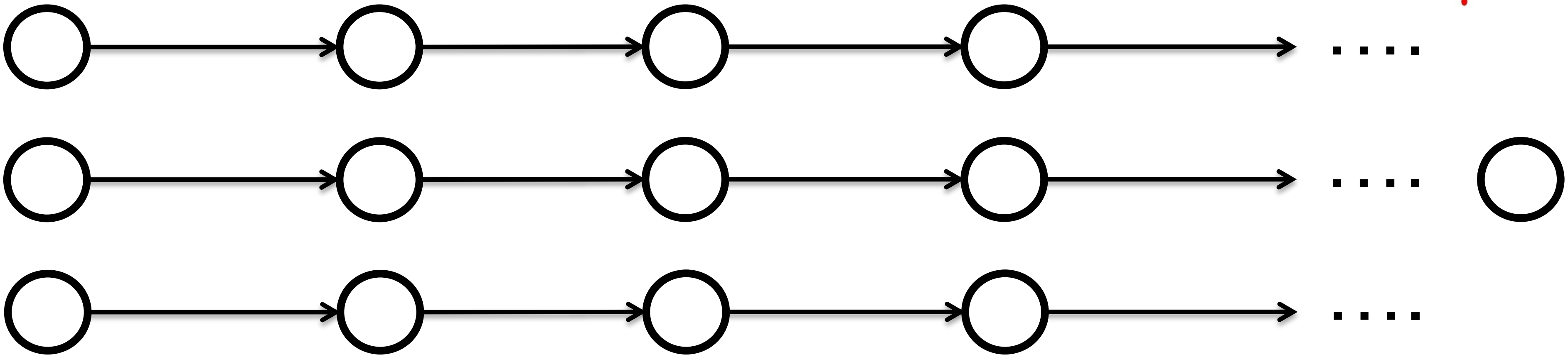
$$
\begin{array}{cc}
 & \begin{array}{ccc} 1 & 0 & \phi \end{array} \\
\begin{array}{c} 1 \\ 0 \\ \phi \end{array} &
\begin{bmatrix}
p_1(1-\mu_{10}) & p_1\mu_{10} & 1-p_1 \\
p_0\mu_{01} & p_0(1-\mu_{01}) & 1-p_0 \\
0 & 0 & 1
\end{bmatrix}
\end{array}
$$

Outline



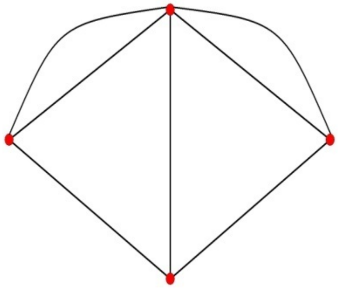
- Model
- Limit learning in long chains
- Optimal regulation/limits on the graph

Multiple Channels



- n(t) chains length t, independent
- As t grows: each less informative, need large n(t)
- $b(t) = P(\omega = 1 \mid \text{information from } n(t) \text{ iid chains})$

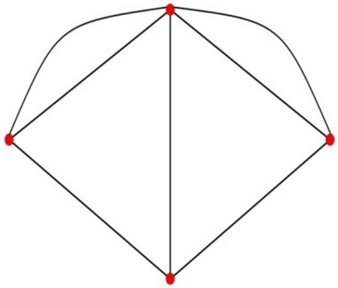
Example



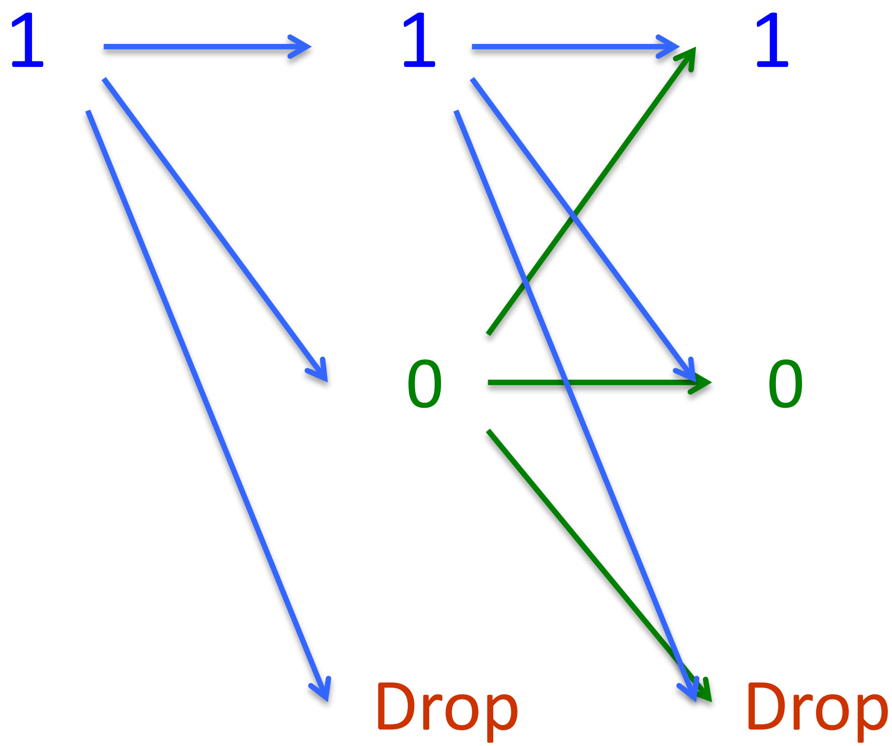
$p=.2 \quad \mu_{01} = .03 \quad \mu_{10} = .10$

1	1	**.18**
	0	**.02**
	Drop	**.80**

Example



$p = .2 \quad \mu_{01} = .03 \quad \mu_{10} = .10$

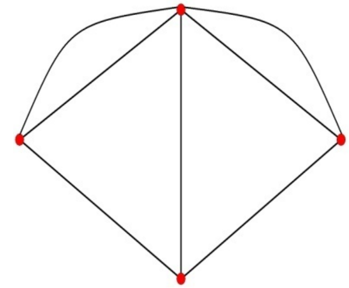


$.2^2\,(.9^2 + .1 \times .03) = .03252$

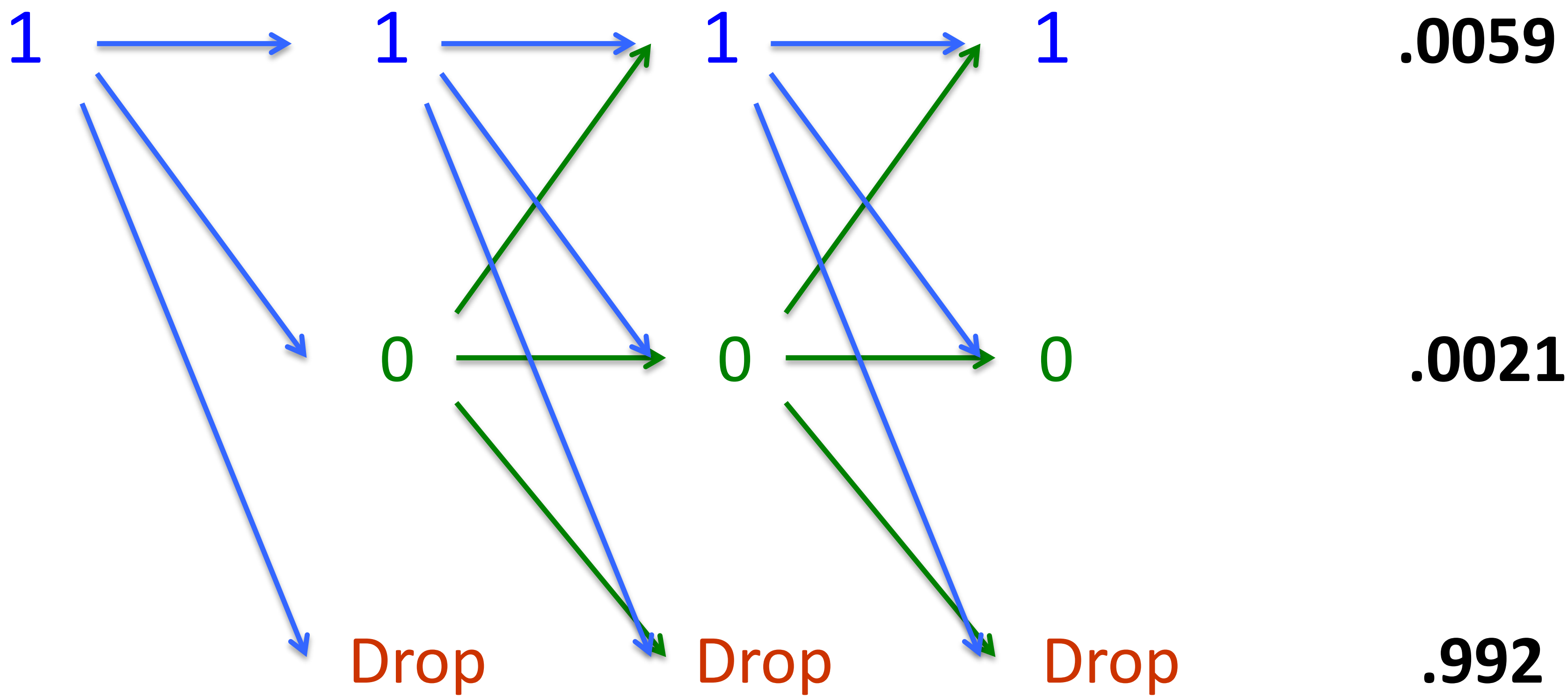
.0075

.96

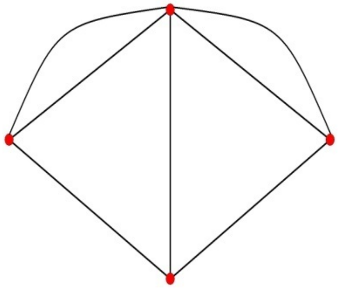
Example



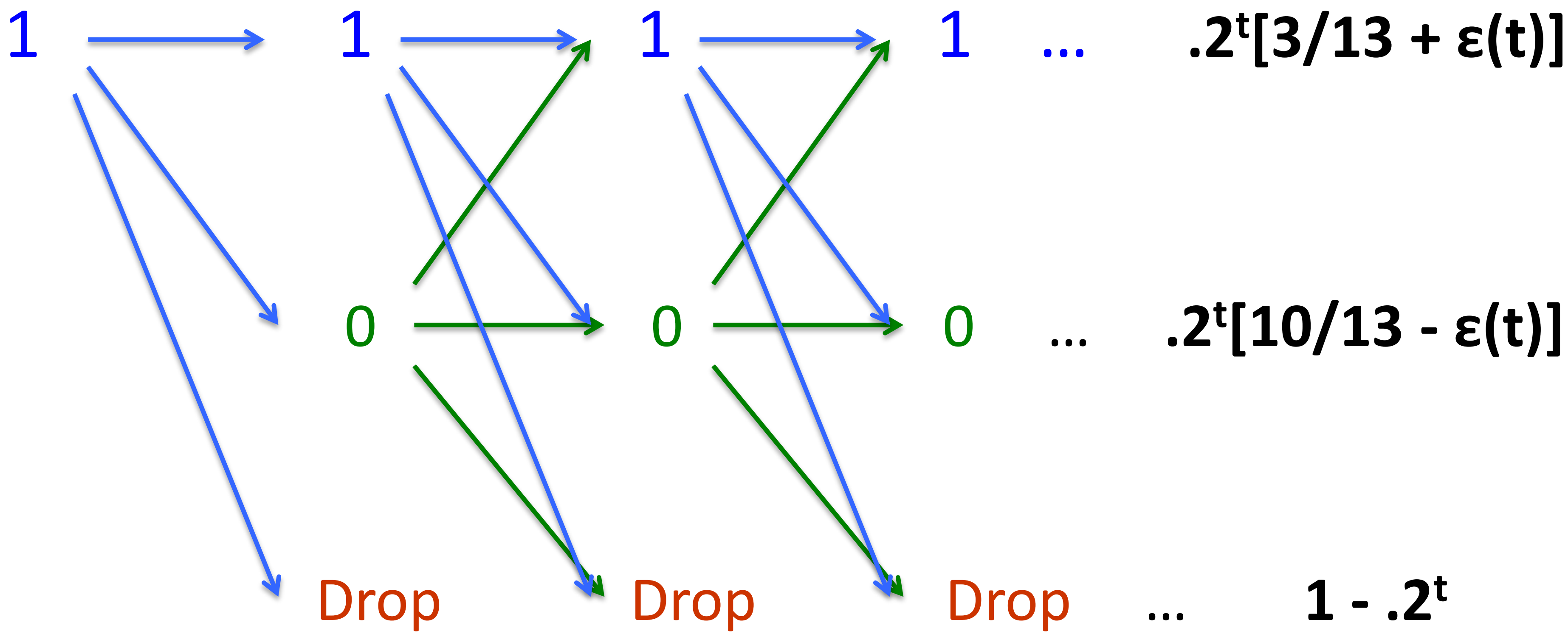
p=.2 μ_{01} =.03 μ_{10} = .10



Example

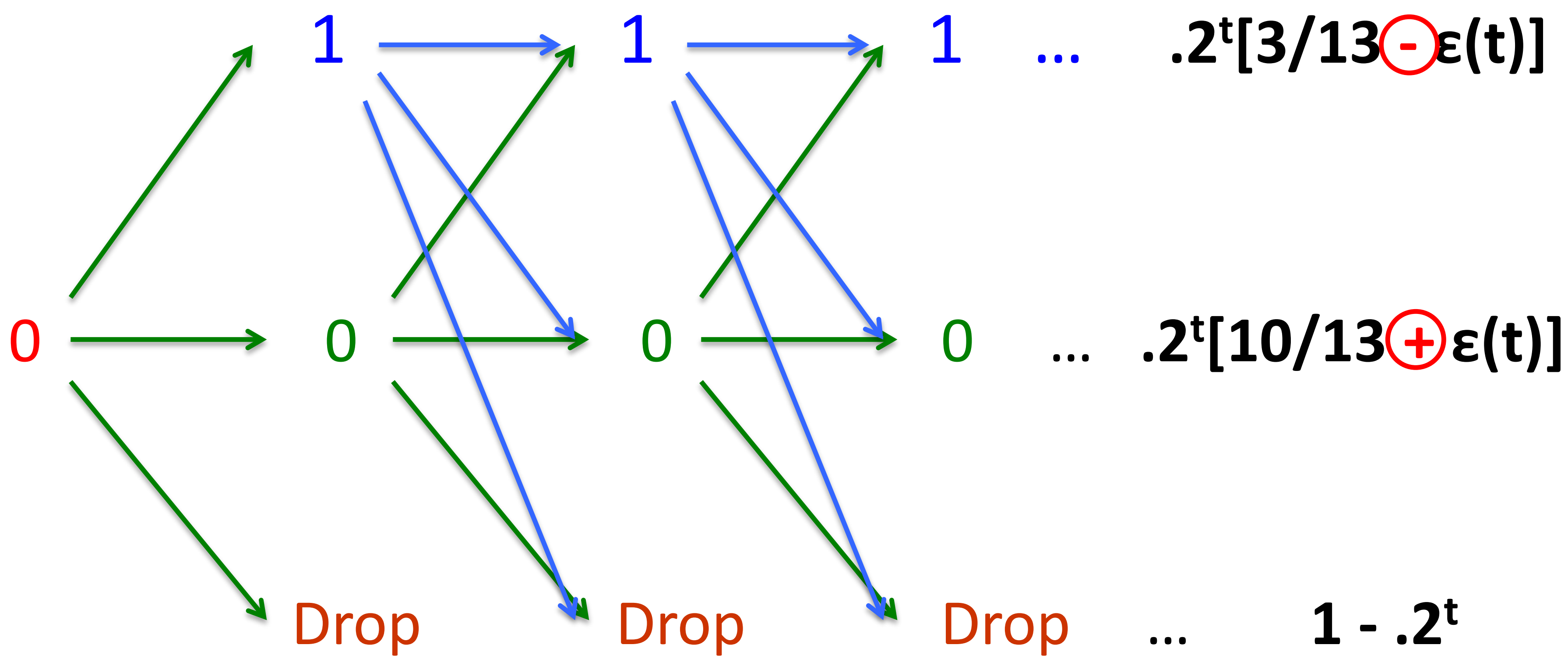


$p = .2 \quad \mu_{01} = .03 \quad \mu_{10} = .10$

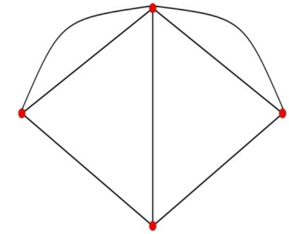


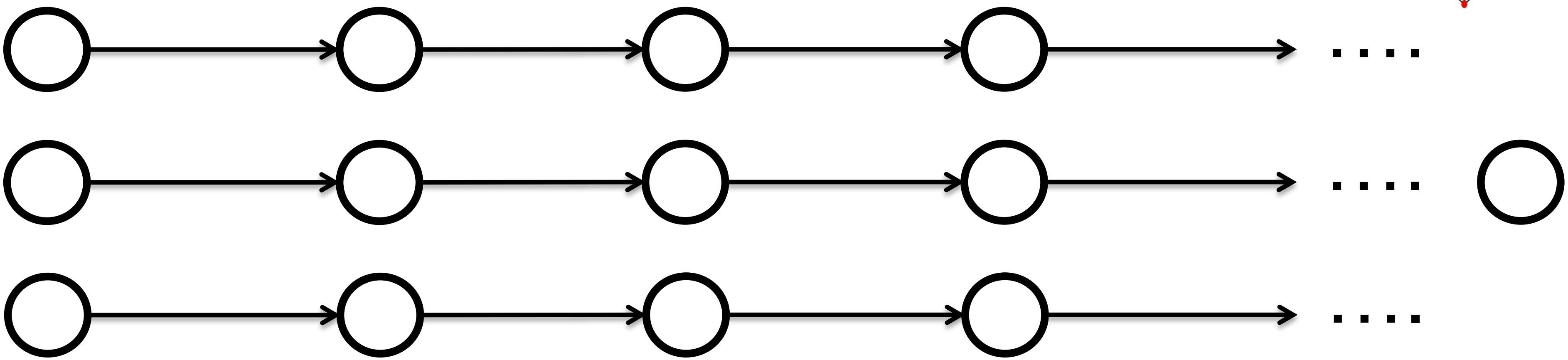
Example

$p=.2 \qquad \mu_{01} = .03 \qquad \mu_{10} = .10$



Learning



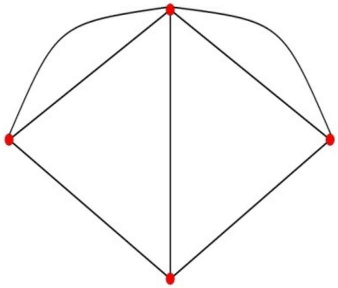


Need many starting chains to overcome:

most messages are dropped

small relative differences of 1/0 depending on starting state after long distances

Learning

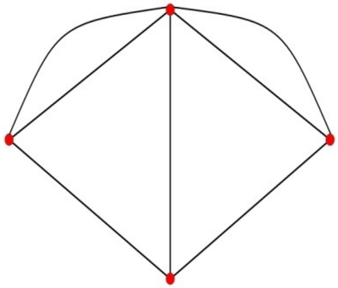


f(t) is a threshold for learning if

- Plim b(t) = 1 or 0 if $n(t)/f(t) \to \infty$
- Plim b(t) = θ if $n(t)/f(t) \to 0$

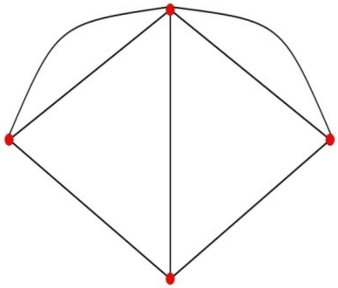
(Note that if lim b(t) = 1 or 0 then beliefs are correct.)

Lemma



If $p > 0$ and μ_{10}, $\mu_{01} < 1/2$ then a threshold for learning is $p^{-t}(1 - \mu_{10} - \mu_{01})^{-2t}$.

Lemma

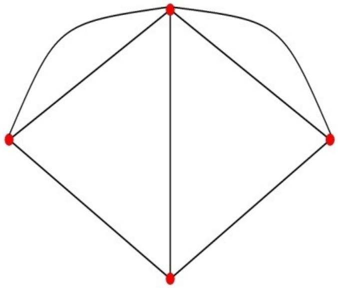


If $p > 0$ and μ_{10}, $\mu_{01} < 1/2$ then a threshold for learning is $p^{-t}(1 - \mu_{10} - \mu_{01})^{-2t}$.

So n(t) needs to grow exponentially in t.

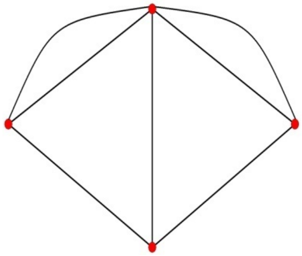
For example, 5^t = thousands, by t=5

Unknown Bias



- μ_{10} and μ_{01} are not known exactly
- Some distribution, possibly very concentrated, but with some interval in its support

Proposition: Impossibility of Learning with Unknown biased agents

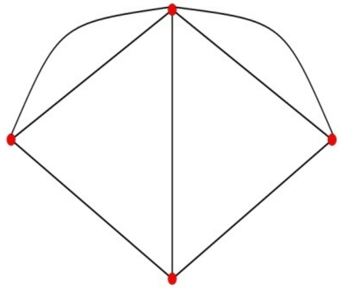


If μ_{10} / μ_{01} is unknown with connected support, then there is imperfect learning:

regardless of $n(t)$ Plim $b(t) = \theta$.

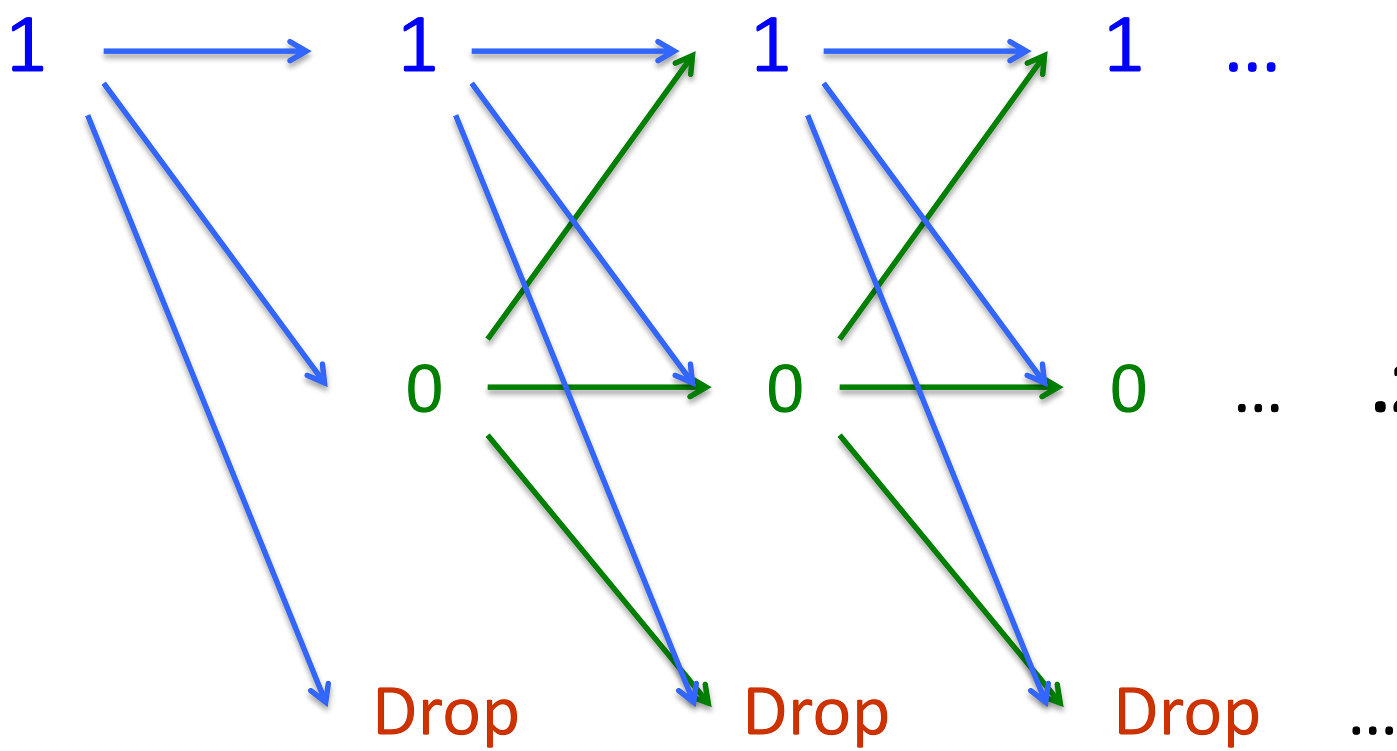
Even small amount of uncertainty about relative bias/mutation rates precludes learning...

Example



$p = .2$ $\cancel{\mu_{01} = .03}$ $\mu_{10} = .10$

$\mu_{01} = .10$



$.2^t[\cancel{3/13} + \varepsilon(t)]$

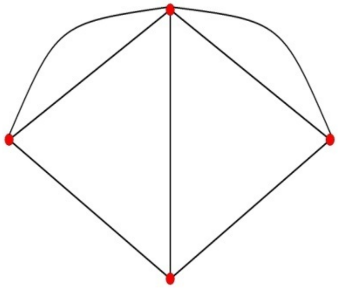
1/2

$.2^t[\cancel{10/13} - \varepsilon(t)]$

1/2

$1 - .2^t$

Intuition

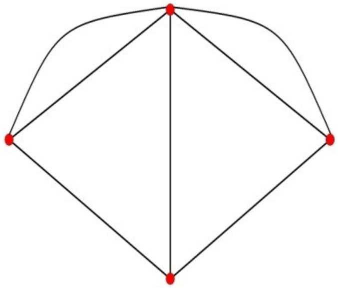


- Limit learning with noise is fragile: learn from *slight* (vanishing) tilt of fraction of signals

- Differences in biases/mutations, non-trivially affects both tilts even in limit.

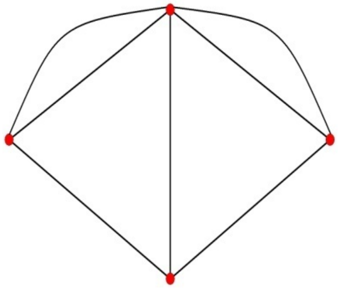
- Even *small* uncertainty about relative probabilities of mutations swamps finer information from starting state

Outline

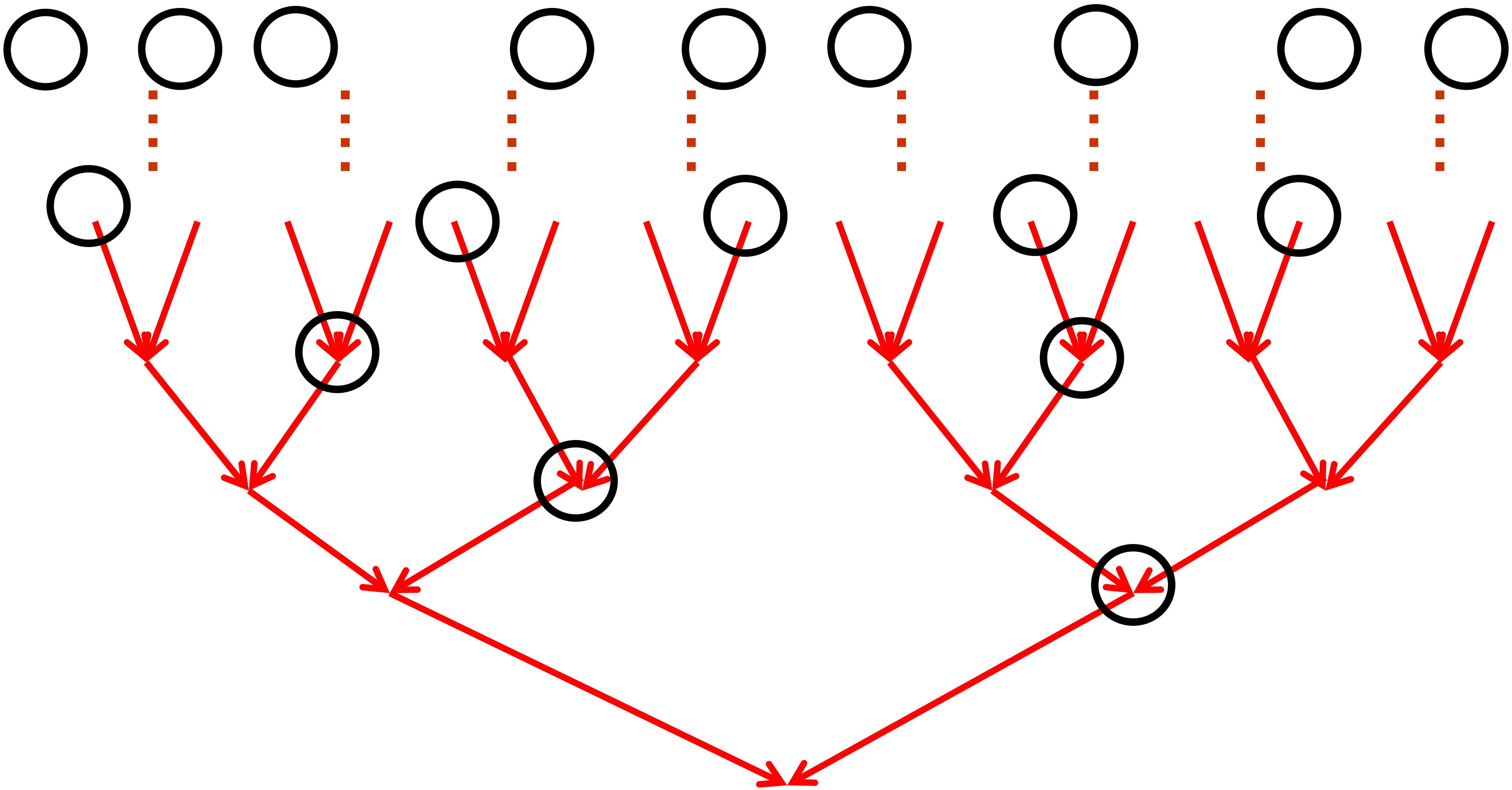


- Model
- Limit learning in long chains
- Optimal regulation/limits on the graph

Partial/*Robust* Learning

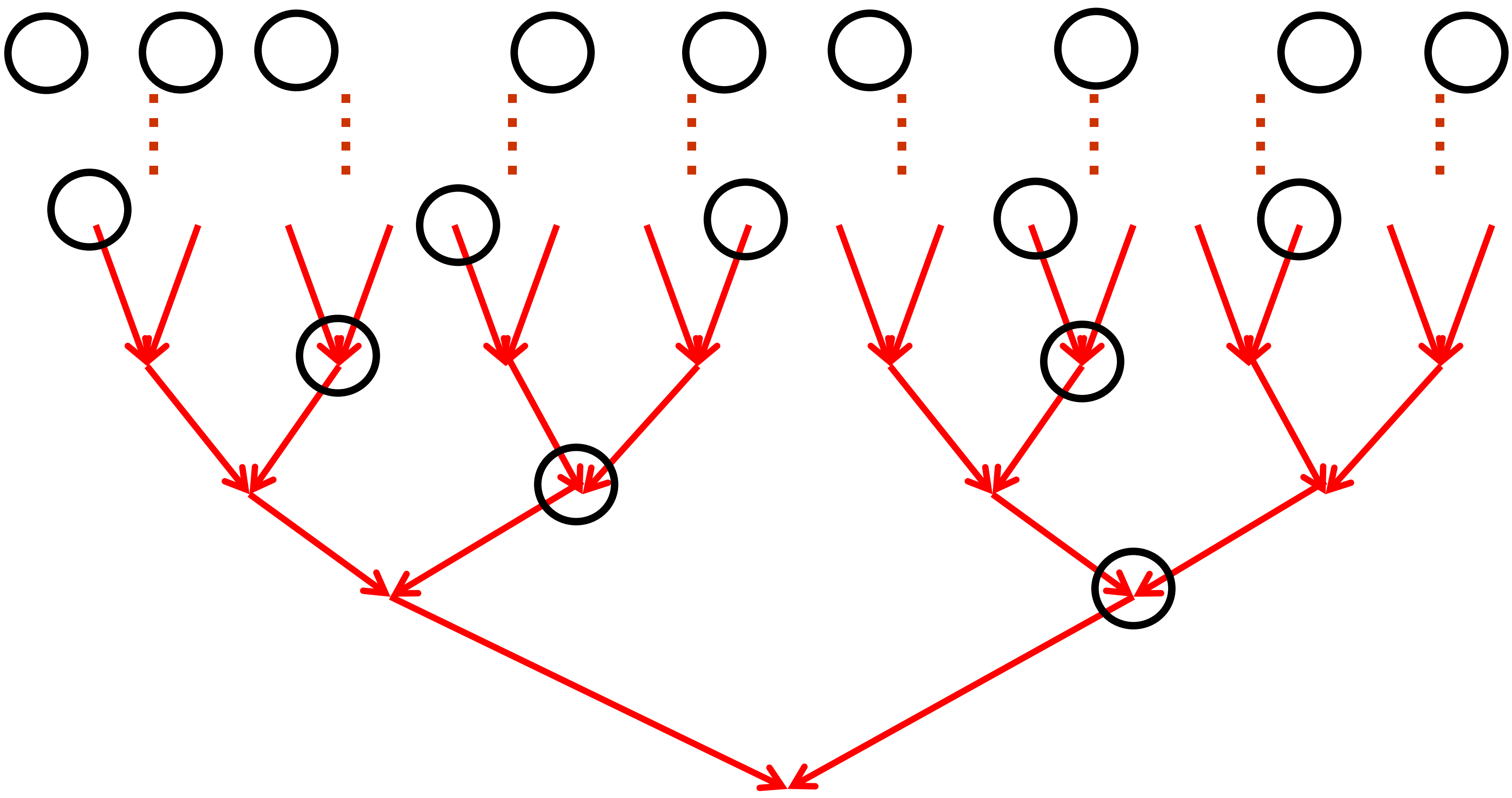


- Learn from near and far
- Without knowing mutation rates and exponentially many sources, distant learning fails
- Although may be far from many sources, can be close to some
- Is partial learning possible then?

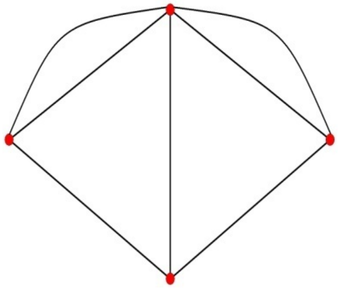


Ratio `far' to `near': $(pd)^{T-t}$

Limit *either* T or pd to decrease this



Trimming the Tree

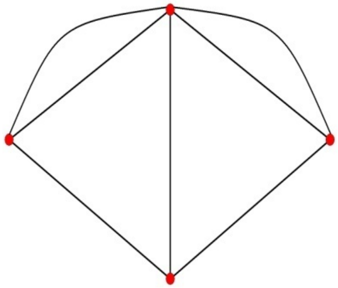


- Information originates: any node, iid probability r
- $p_0 = p_1 = p$
- μ_{10} and μ_{01} can differ and are unknown
- Take receiver to have symmetric prior over state and μ_{10} and μ_{01}:

  Go with the majority of signals

Robust learning: majority will be correct, in expectation

Partial/*Robust* Learning

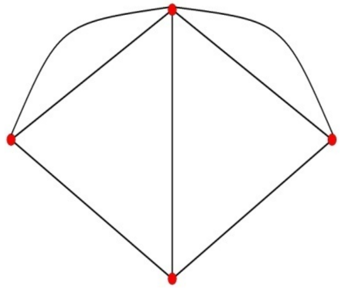


- Learn from near and far, *improve ratio near/far*

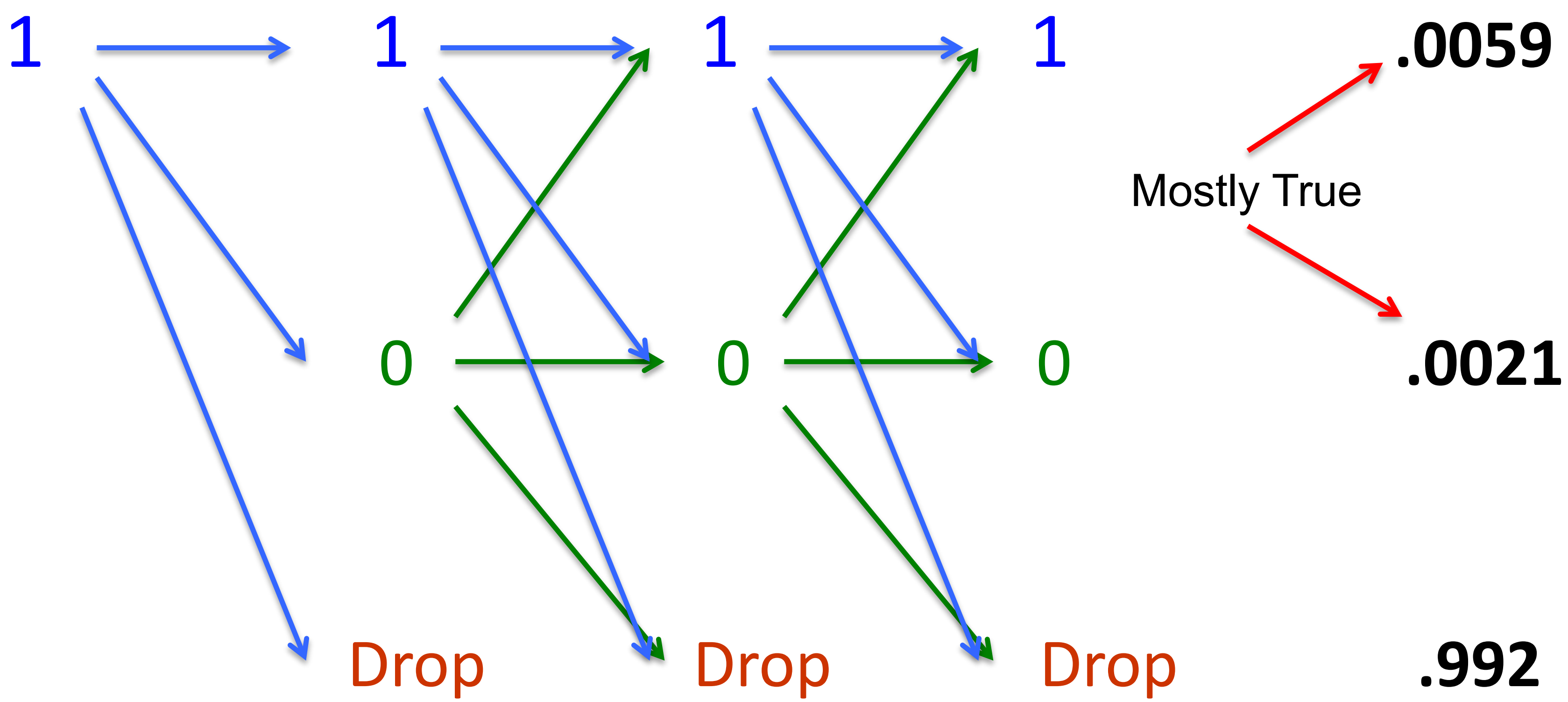
- Trim the tree:
  - Limit depth
  - Limit breadth

- Consider G-W random tree with average degree d and some depth T

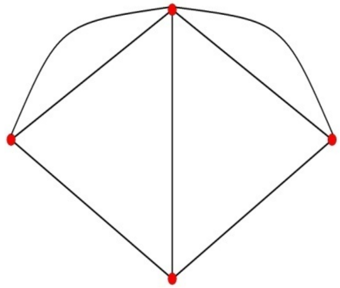
Example



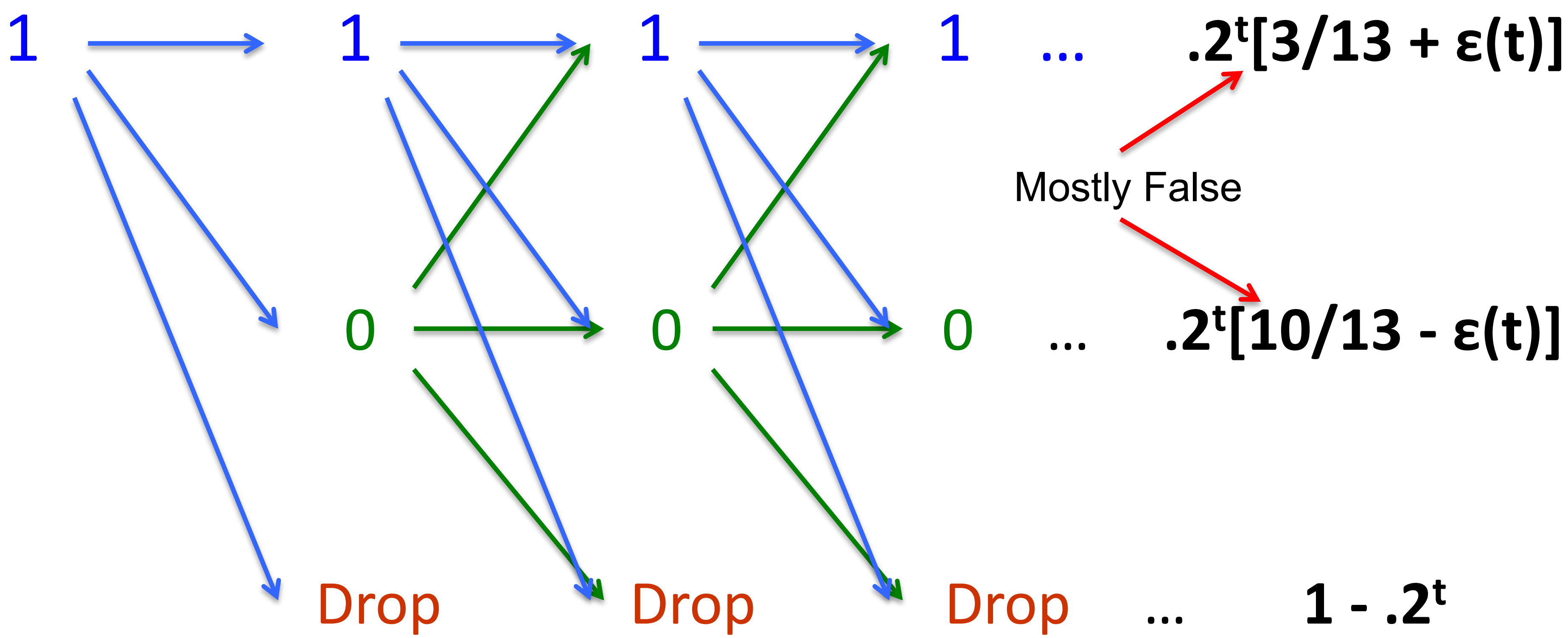
$p=.2 \quad \mu_{01}=.03 \quad \mu_{10}=.10$



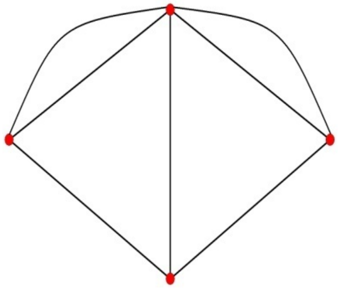
Example



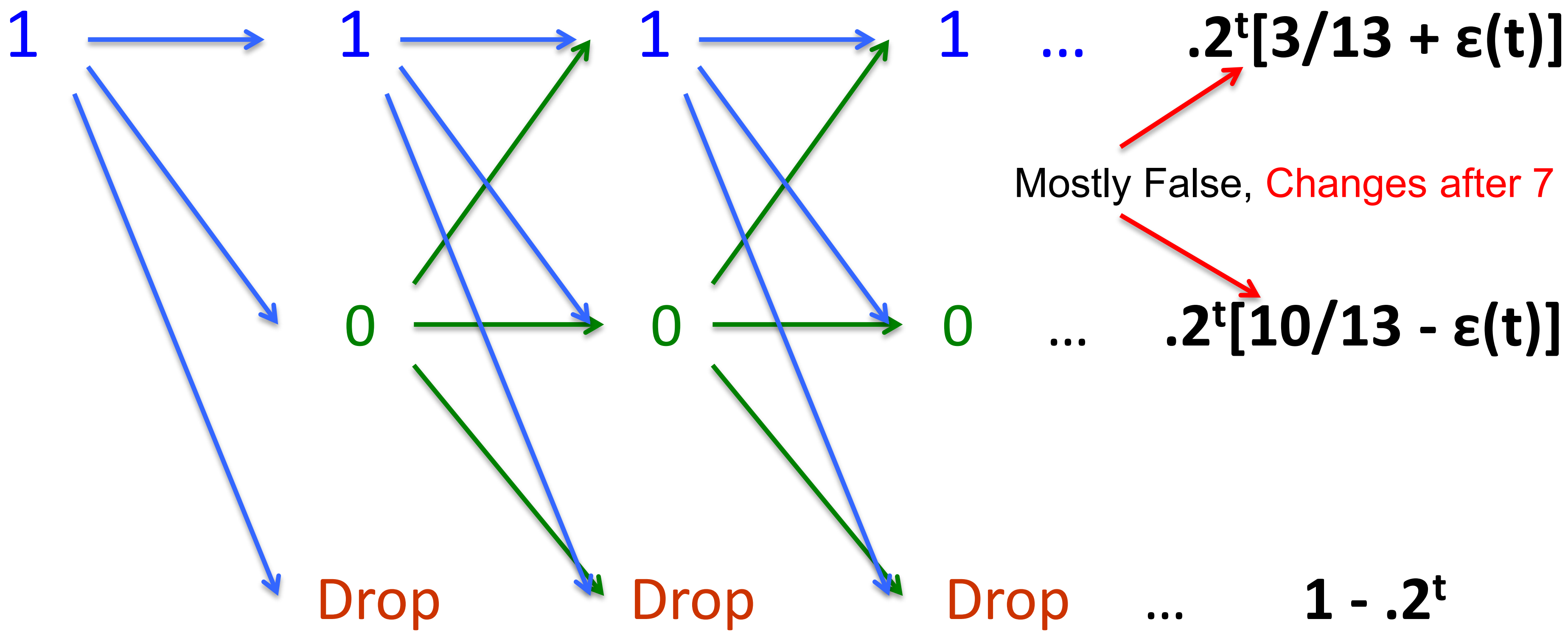
$p = .2 \quad \mu_{01} = .03 \quad \mu_{10} = .10$



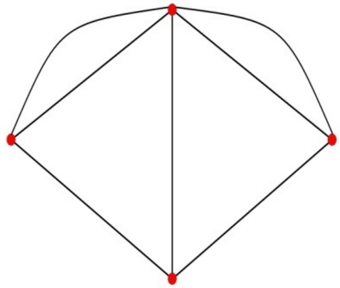
Example



$p=.2 \quad \mu_{01} = .03 \quad \mu_{10} = .10$



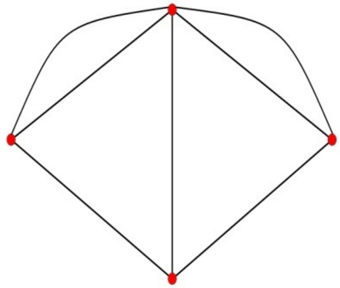
Expected to be correct:



Cap messages to travel distance no longer than

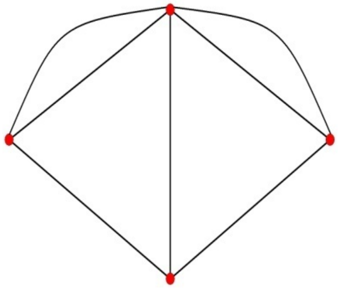
$$T = \frac{\log(1/2 - \min[\mu_{01}/(2\mu_{10}), \mu_{10}/(2\mu_{01})])}{\log(1 - \mu_{01} - \mu_{10})}$$

Example



- p =.2 μ_{10} = .10 and μ_{01} = .03 - so bias is towards 0
- If true state is 0, no problem – most messages are true regardless of network or distance travelled
- If true state is 1, then messages that travel too far are more likely false
- T = 7.5 is the threshold

Trimming Breadth



- If cannot cap distance, then try to ensure that more messages come from close than far: cap d

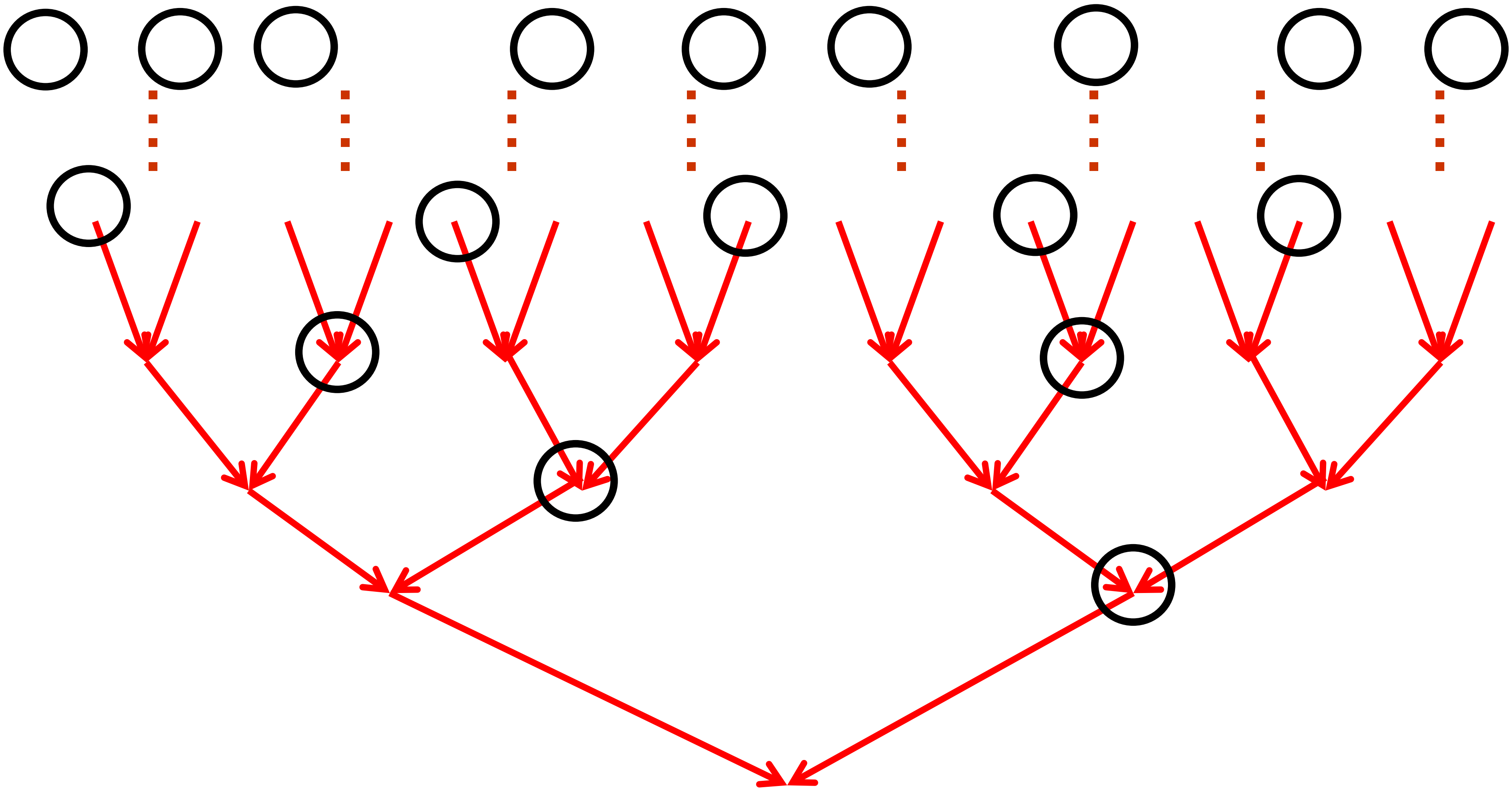
- As breadth of tree increases

  – exponentially increases the number of nodes at further distances compared to closer

  – cutting breadth improves ratio near to far

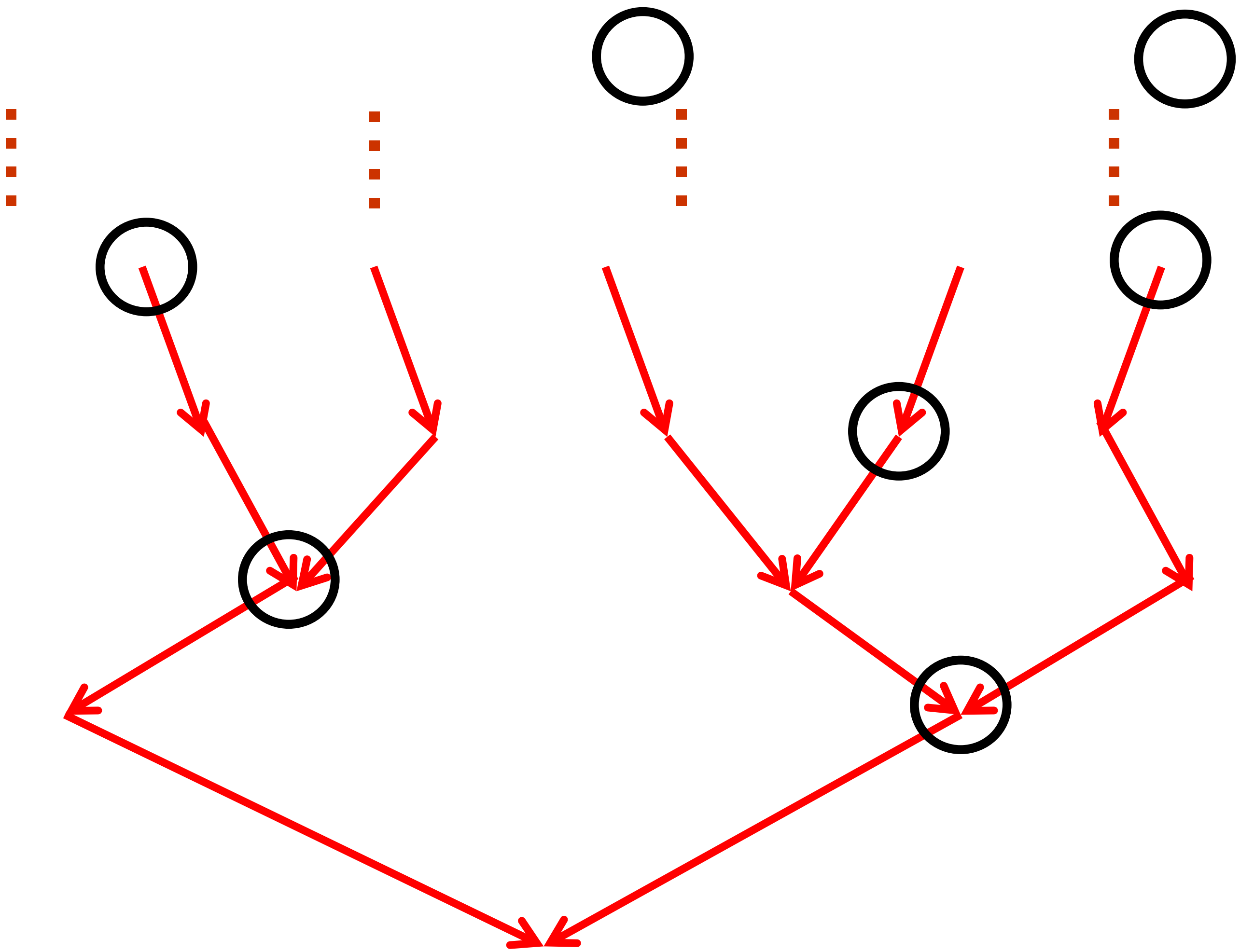
Ratio `far' to `near': $(pd)^{T-t}$

Limit *either* T or pd to decrease this

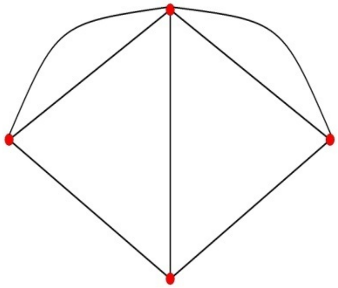


Ratio `far’ to `near’: $(pd)^{T-t}$

Limit *either* T or pd to decrease this



Propostion: Majority True

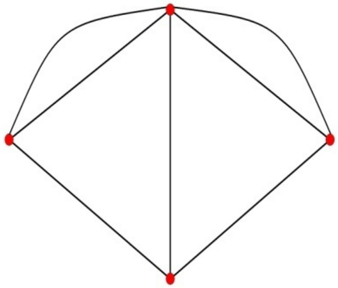


T unbounded. Galton-Watson process.

The expected number of true messages exceeds the expected number of false messages in both states *if and only if*

$$pd < [1 - 2\max\{\mu_{01}, \mu_{10}\}] / (1 - \mu_{01} - \mu_{10})$$

Propostion: Majority True

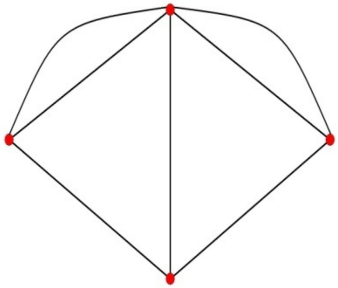


$pd < [1 - 2\max\{\mu_{01}, \mu_{10}\}] / (1 - \mu_{01} - \mu_{10})$

Cap is decreasing in p – messages travel further, more likely to mutate

decreasing in max mutation rate, increasing in the min rate,

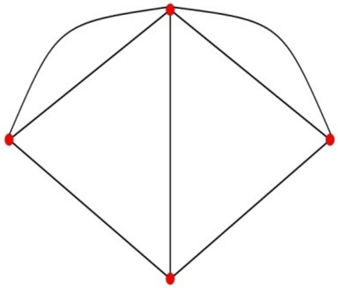
Caps



- With $\mu_{01} = .03$, $\mu_{10} = .10$, $p = .20$,
- Cap on T is 7.5
- If cannot cap T, then cap on d is 4.6

(WhatsApp, Facebook recently capped d at 5)

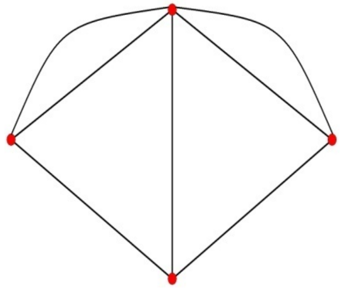
Breadth Cap



If μ_{01} and μ_{10} differ by a lot, and p is large, then it is *impossible* to satisfy…

many messages then come from far, and are very likely to be false in the state not favored by mutations

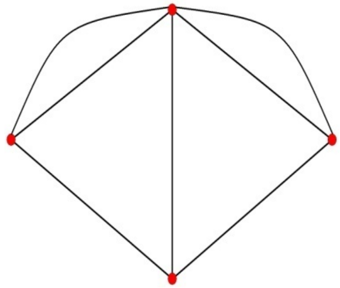
Optimum



Cap ensures that there are more correct than incorrect

What if we want to maximize that difference?

Proposition: Majority True



T unbounded. Take $\mu_{01} < \mu_{10}$

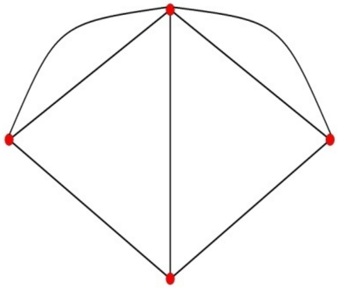
The expected number of true messages minus false messages is maximized at

$$d^* = [1 - Z]/ [p(1-(1 - \mu_{01} - \mu_{10})Z)]$$

$$\text{where } Z = [\mu_{10} - \mu_{01}]^{.5}/[2\,\mu_{10}(1 - \mu_{01} - \mu_{10})]^{.5}$$

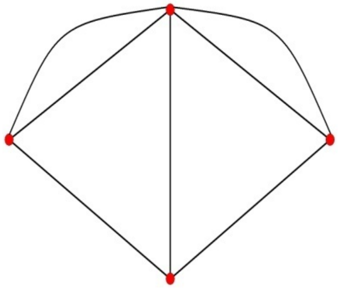
Similar comparative statics to cap expression

Caps, Optima



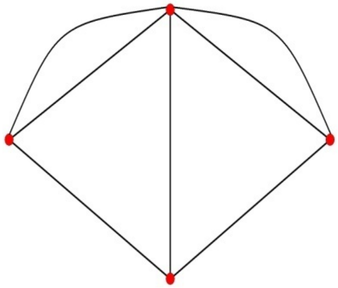
- With $\mu_{01} = .03$, $\mu_{10} = .10$, $p = .20$,
- Cap on T is 7.5
- Cap on d is 4.6 if $p = .2$, 9.2 if $p = .1$, 18.4 if $p = .05$
- Maximizing d is 1.8 if $p = .2$, 3.6 if $p = .1$, 7.2 if $p = .05$
- WhatsApp, Facebook recently capped d at 5

Summary



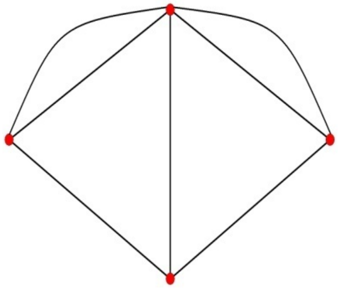
- Noise: mutating, dropping, biased sending
- Need lots of sources to learn
- Slight uncertainty in mutation rates kills full learning
- *Crude, simple caps on passing can make partial learning possible*

Thoughts



- (Some) networks work well at aggregating info, when not noisy
- But, they face significant challenges when noisy
- Shorter chains help
- One of the drivers of polarization? Only trust nearby information?
- Patterns of mis/disinformation can matter…

Thank you!



Extra Slides