Machine Learning for Data Science

Spring 2015

Lecture 26:
Punchlines and morals

http://upload.wikimedia.org/wikipedia/commons/5/56/Alice_par_John_Tenniel_32.png

The inspiration for this lecture

[link to fairy tale in True Facts About the Owl, by zefrank1]

What is the moral of our story?

[The Duchess said,] `You're thinking about something, my dear, and that makes you forget to talk. I can't tell you just now what the moral of that is, but I shall remember it in a bit.'

`Perhaps it hasn't one,' Alice ventured to remark.

`Tut, tut, child!' said the Duchess. `Everything's got a moral, if only you can find it.'

http://upload.wikimedia.org/wikipedia/commons/5/56/Alice_par_John_Tenniel_32.png
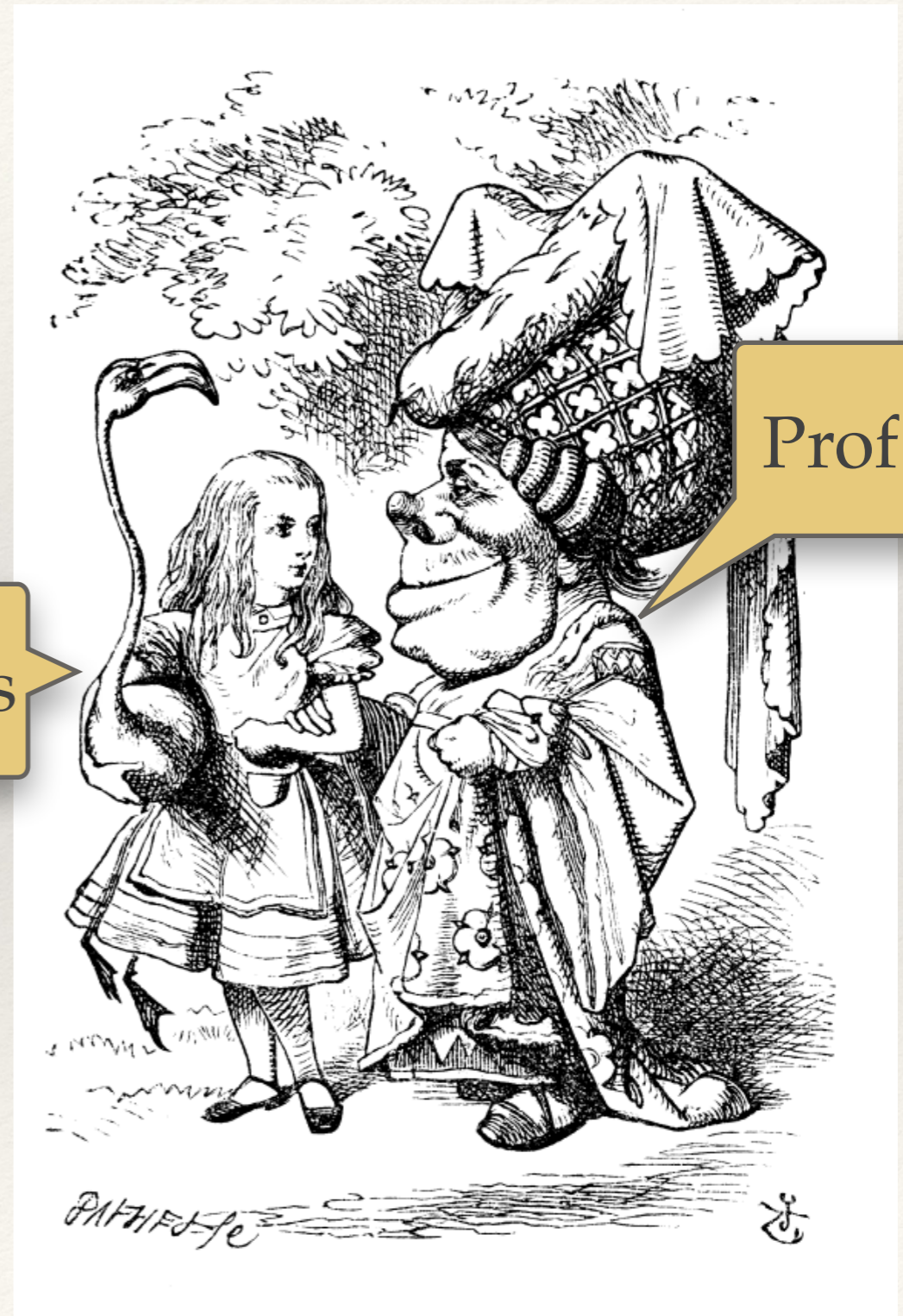
What is the moral of our story?

[The Duchess said,] `You're thinking about something, my dear, and that makes you forget to talk. I can't tell you just now what the moral of that is, but I shall remember it in a bit.'

`Perhaps it hasn't one,' Alice ventured to remark.

`Tut, tut, child!' said the Duchess. `Everything's got a moral, if only you can find it.'

Prof Lee

What is the moral of our story?

[The Duchess said,] `You're thinking about something, my dear, and that makes you forget to talk. I can't tell you just now what the moral of that is, but I shall remember it in a bit.'

`Perhaps it hasn't one,' Alice ventured to remark.

`Tut, tut, child!' said the Duchess. `Everything's got a moral, if only you can find it.'

Prof Lee

you guys

What is the moral of our story?

[The Duchess said,] `You're thinking about something, my dear, and that makes you forget to talk. I can't tell you just now what the moral of that is, but I shall remember it in a bit.'

`Perhaps it hasn't one,' Alice ventured to remark.

`Tut, tut, child!' said the Duchess. `Everything's got a moral, if only you can find it.'

Prof Sridharan

Prof Lee

you guys

http://upload.wikimedia.org/wikipedia/commons/5/56/Alice_par_Joh

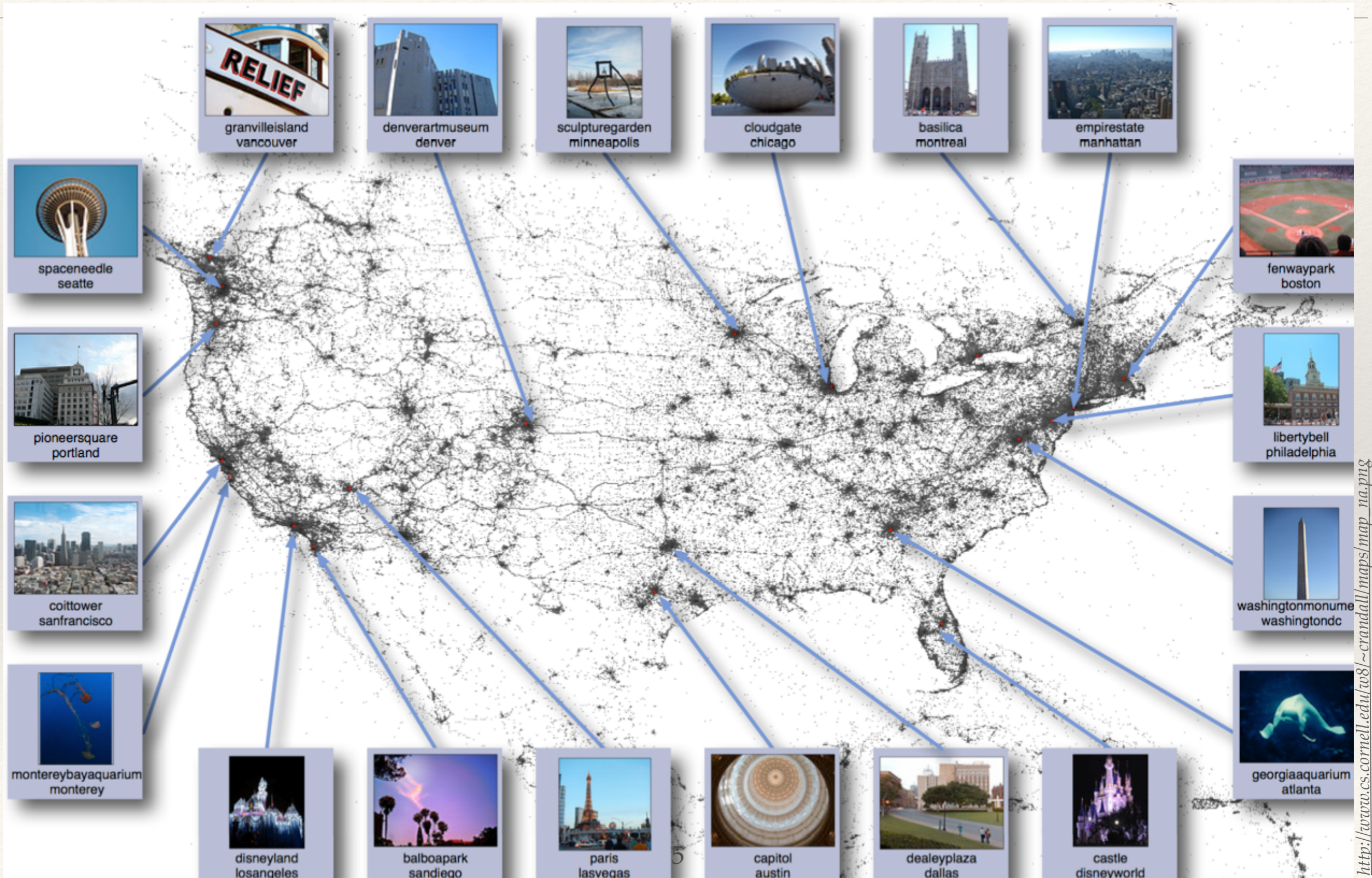# Data relationships are key

- Between features (as columns) - dimension reduction

- Between items - clustering

- Between variables - graphical models

Example: Learning from geotagged, captioned image data

(Crandall, Backstrom, Huttenlocher, Kleinberg, best paper nominee, WWW 2009)

But there is no free lunch

❖ Complexity of algorithm, data requirements, need for prior knowledge, ….

Seek a "good" feature space

Cholera deaths (examples based off Edmund Cooper, 1854):

- ❖ 40 Broad Street: next to gully-hole, drain in good condition, …

- ❖ 39 Broadwick St: not next to gully-hole, serviced by new sewer, …

- ❖ …

Brody, Rip, Vinten-Johansen, Paneth, and Rachman "Map-making and myth-making in Broad Street: the London cholera epidemic, 1854", *The Lancet* (356):64-68, 2000.

Seek a "good" feature space

Cholera deaths (examples based off Edmund Cooper, 1854):

❖ 40 Broad Street: next to gully-hole, drain in good condition, …

❖ 39 Broadwick St: not next to gully-hole, serviced by new sewer, …

❖ …

Map by John Snow
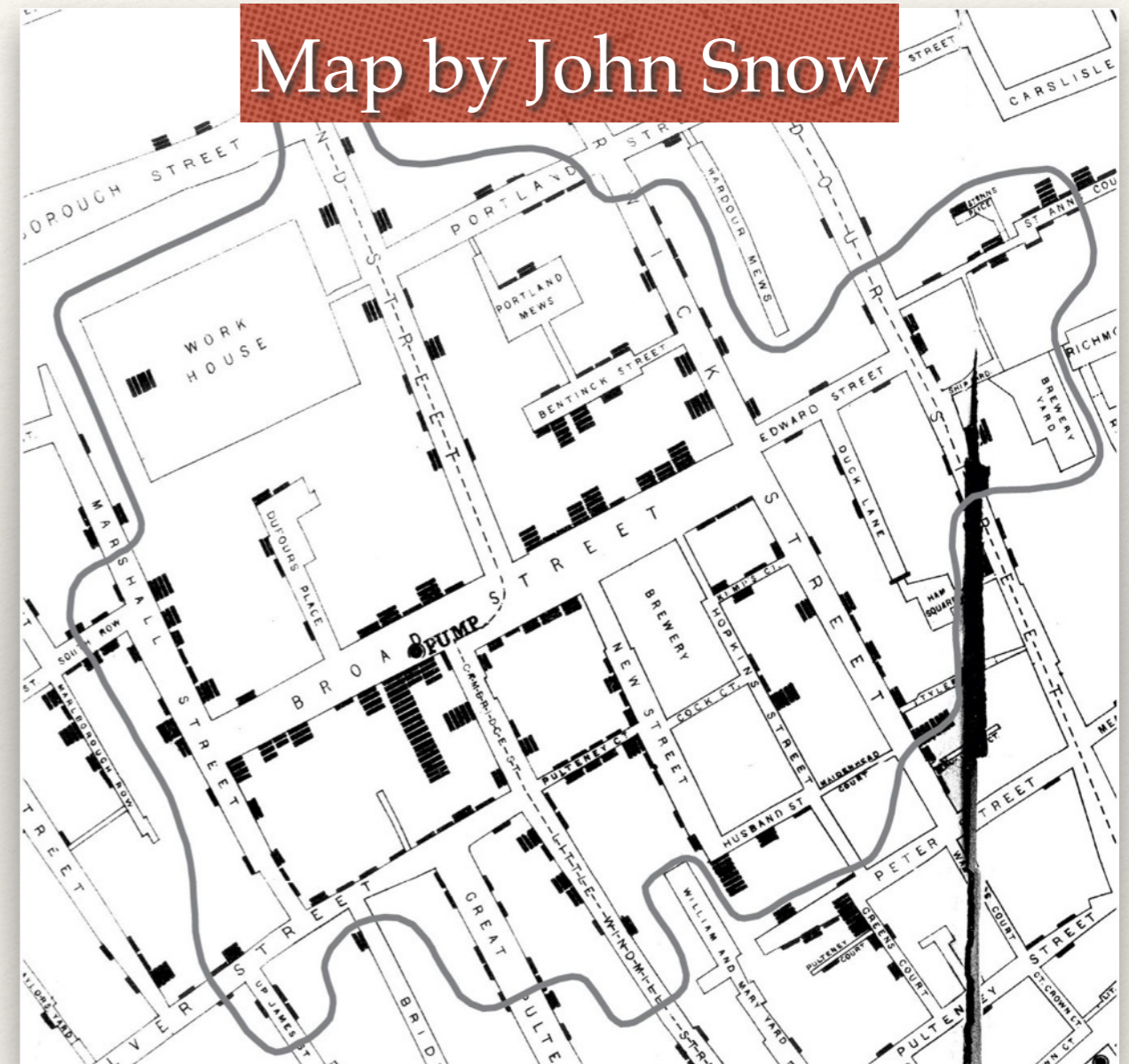
http://johnsnow.matrix.msu.edu/images/online_companion/chapter_images/fig12-6.jpg

Brody, Rip, Vinten-Johansen, Paneth, and Rachman "Map-making and myth-making in Broad Street: the London cholera epidemic, 1854", *The Lancet* (356):64-68, 2000.

Watch for outliers (bad news)
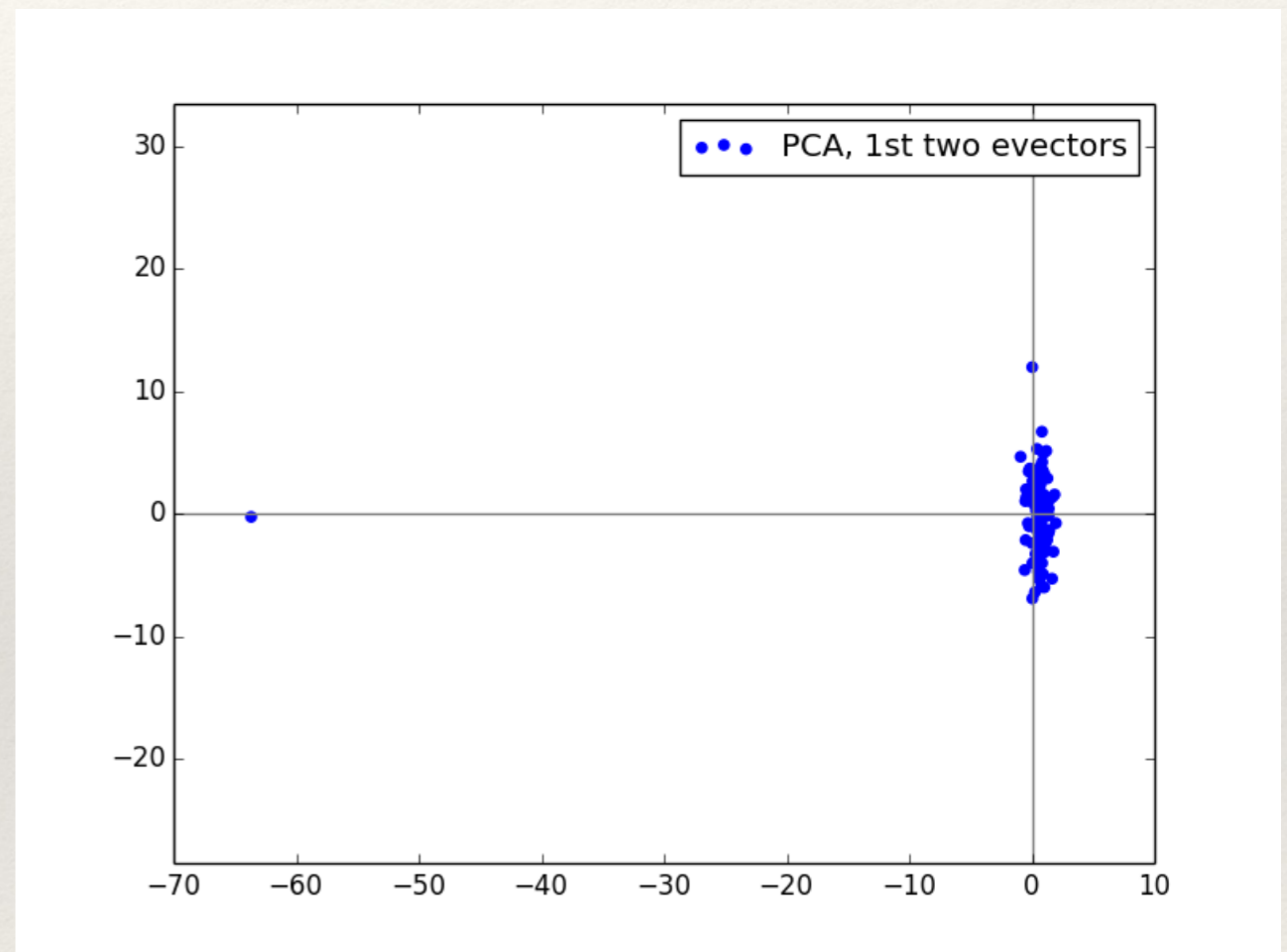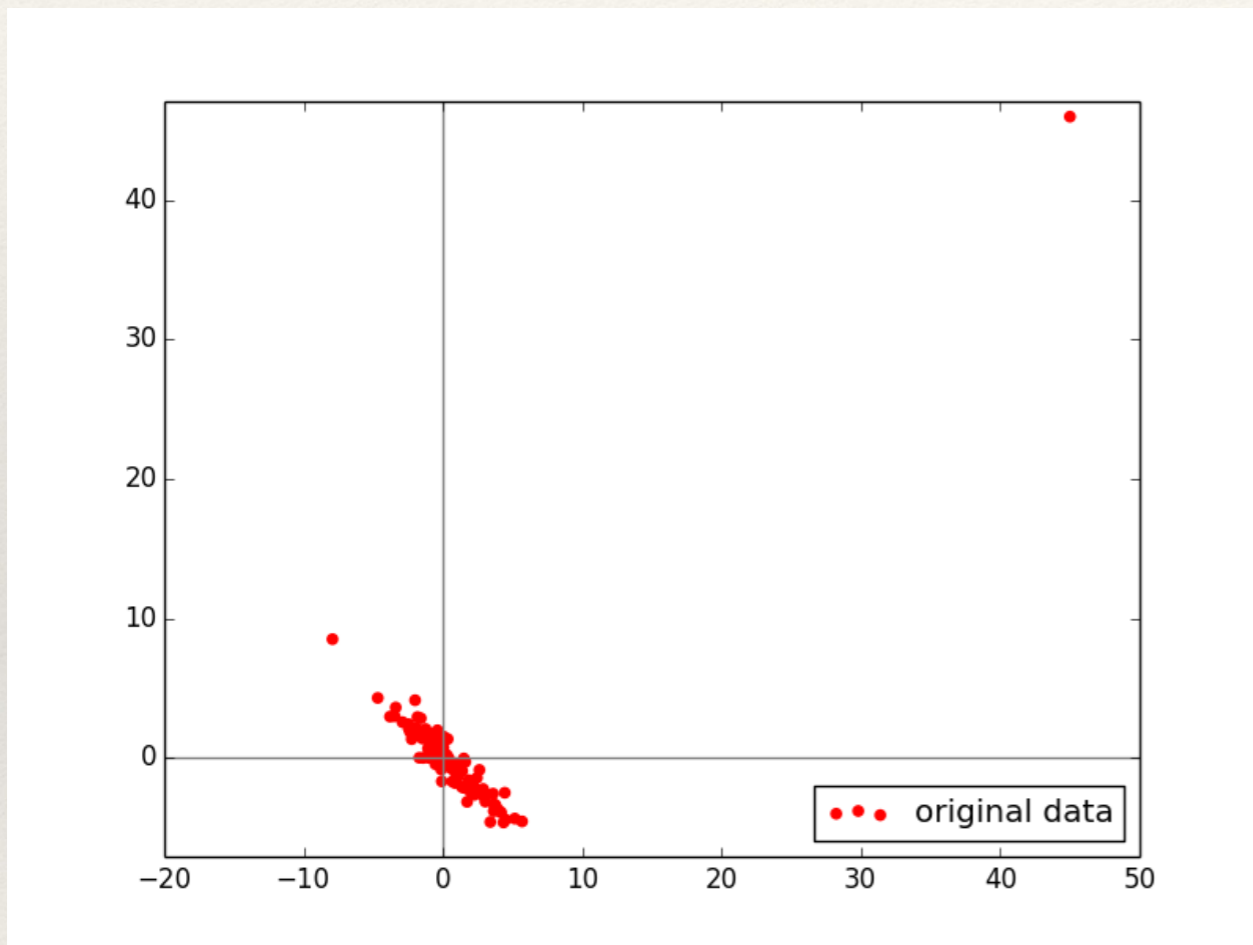
❖ Similar issues with clustering

100% is not the goal

❖ Don't get the wrong idea from the two course competitions!

Fig. 4. Over-enthusiastic data fitters

100% is not the goal

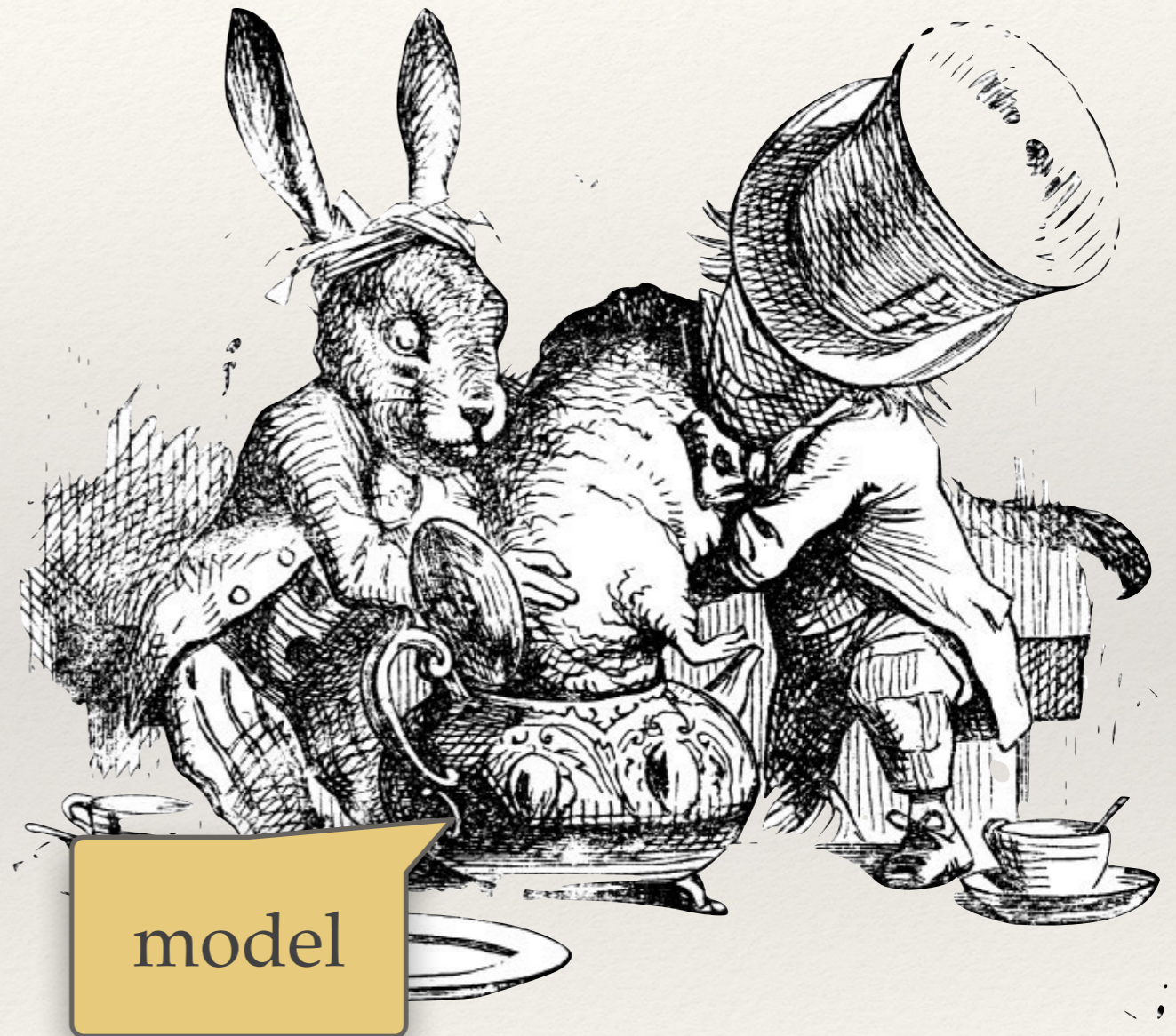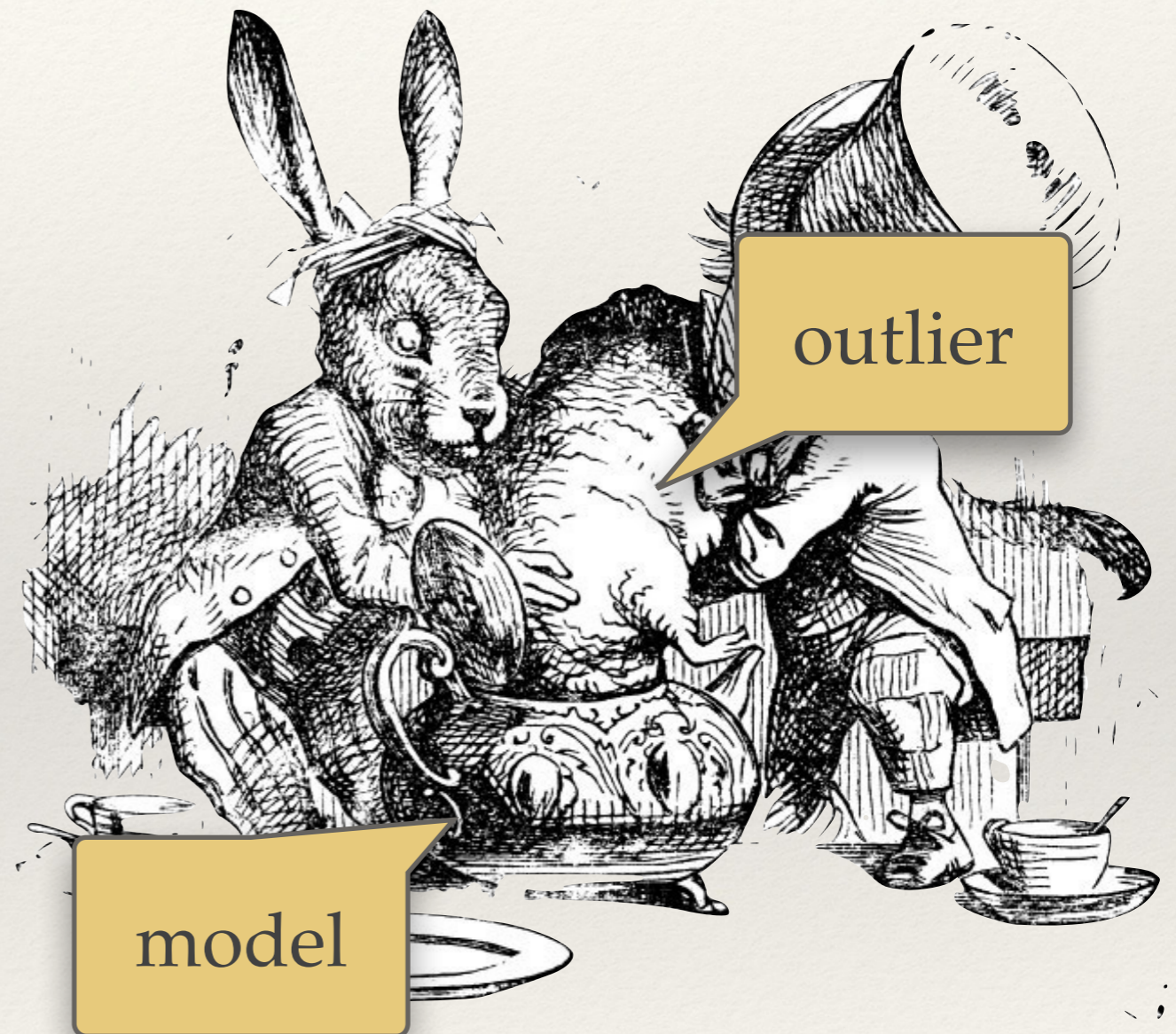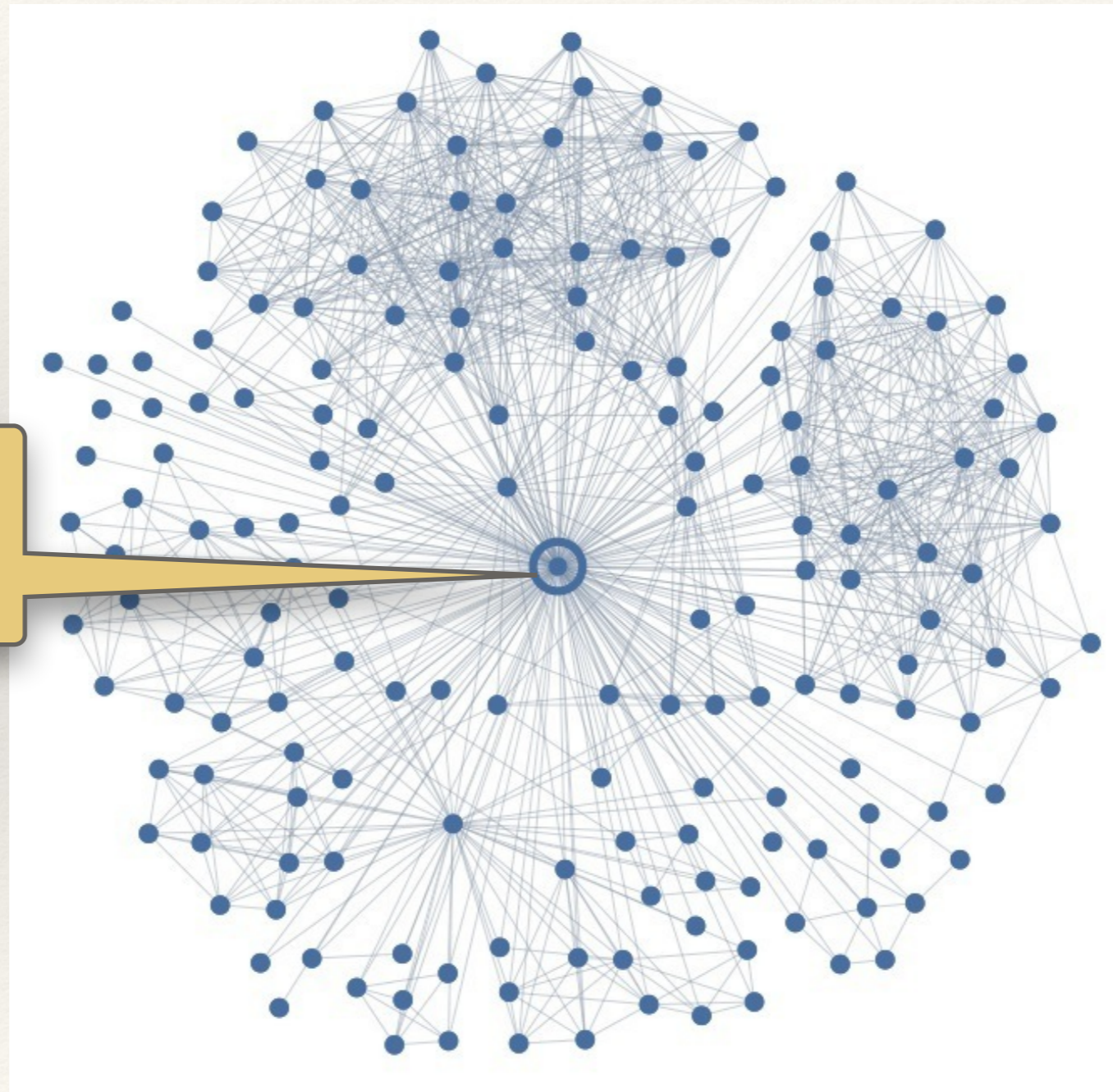❖ Don't get the wrong idea from the two course competitions!

model

Fig. 4. Over-enthusiastic data fitters

100% is not the goal

❖ Don't get the wrong idea from the two course competitions!

outlier

model

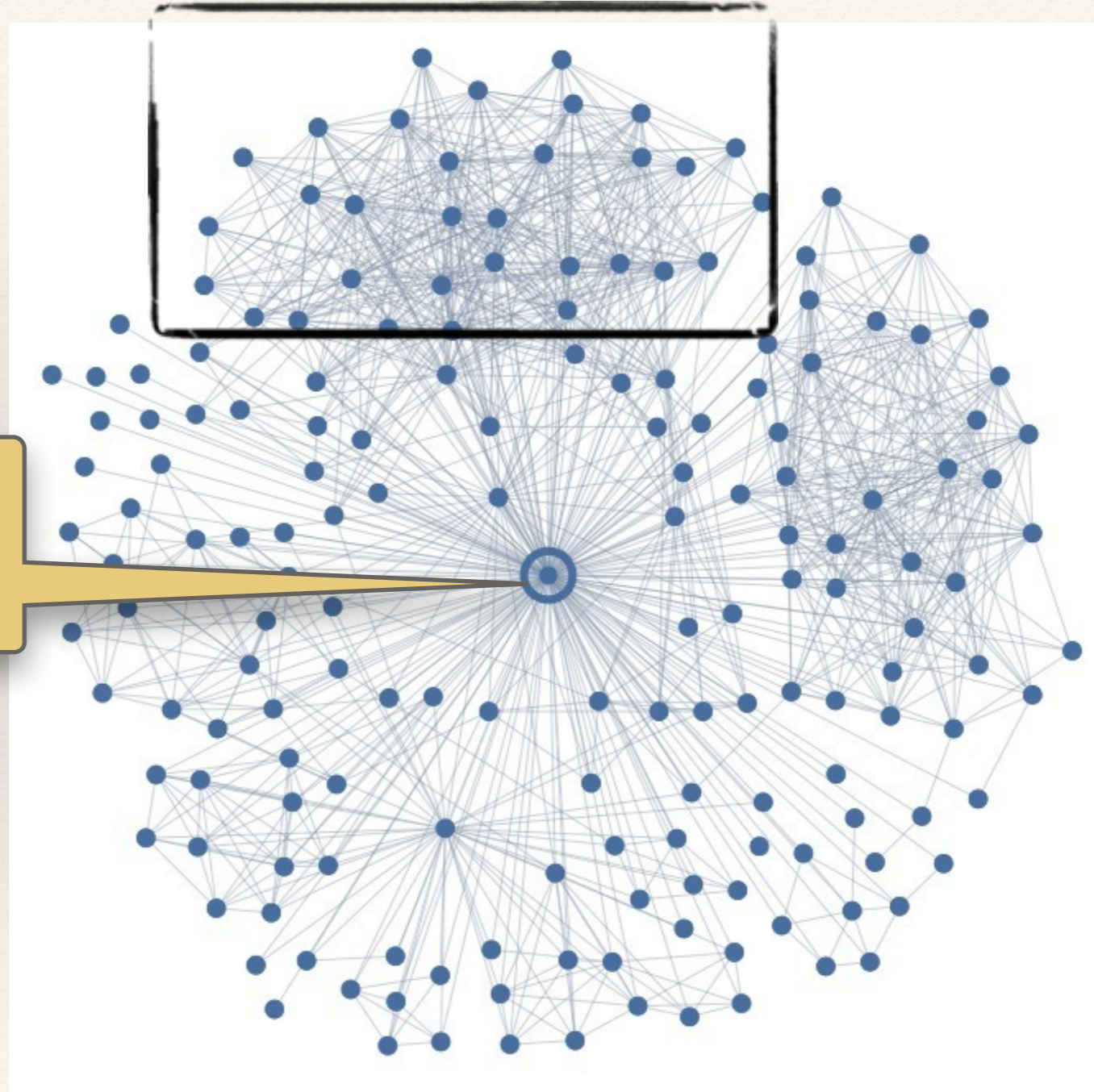Fig. 4. Over-enthusiastic data fitters

Watch for outliers (good news)

Somebody on Facebook

http://overstated.net/wp/uploads/2009/03/asmith-connections.png

10

Watch for outliers (good news)

Somebody on Facebook

10

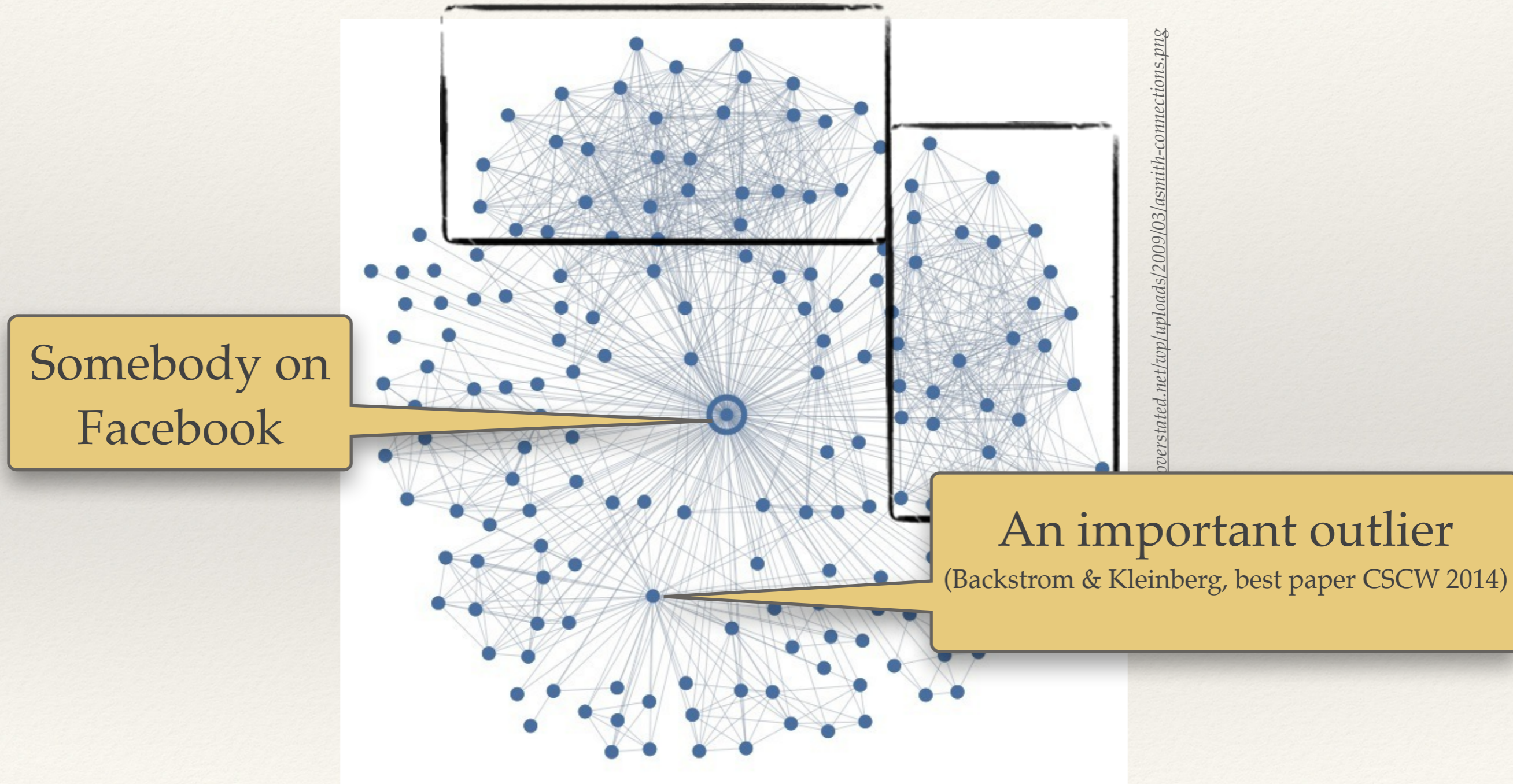# Watch for outliers (good news)

Somebody on Facebook

http://overstated.net/wp/uploads/2009/03/asmith-connections.png

Watch for outliers (good news)

Somebody on Facebook

An important outlier
(Backstrom & Kleinberg, best paper CSCW 2014)

overstated.net/wp/uploads/2009/03/asmith-connections.png
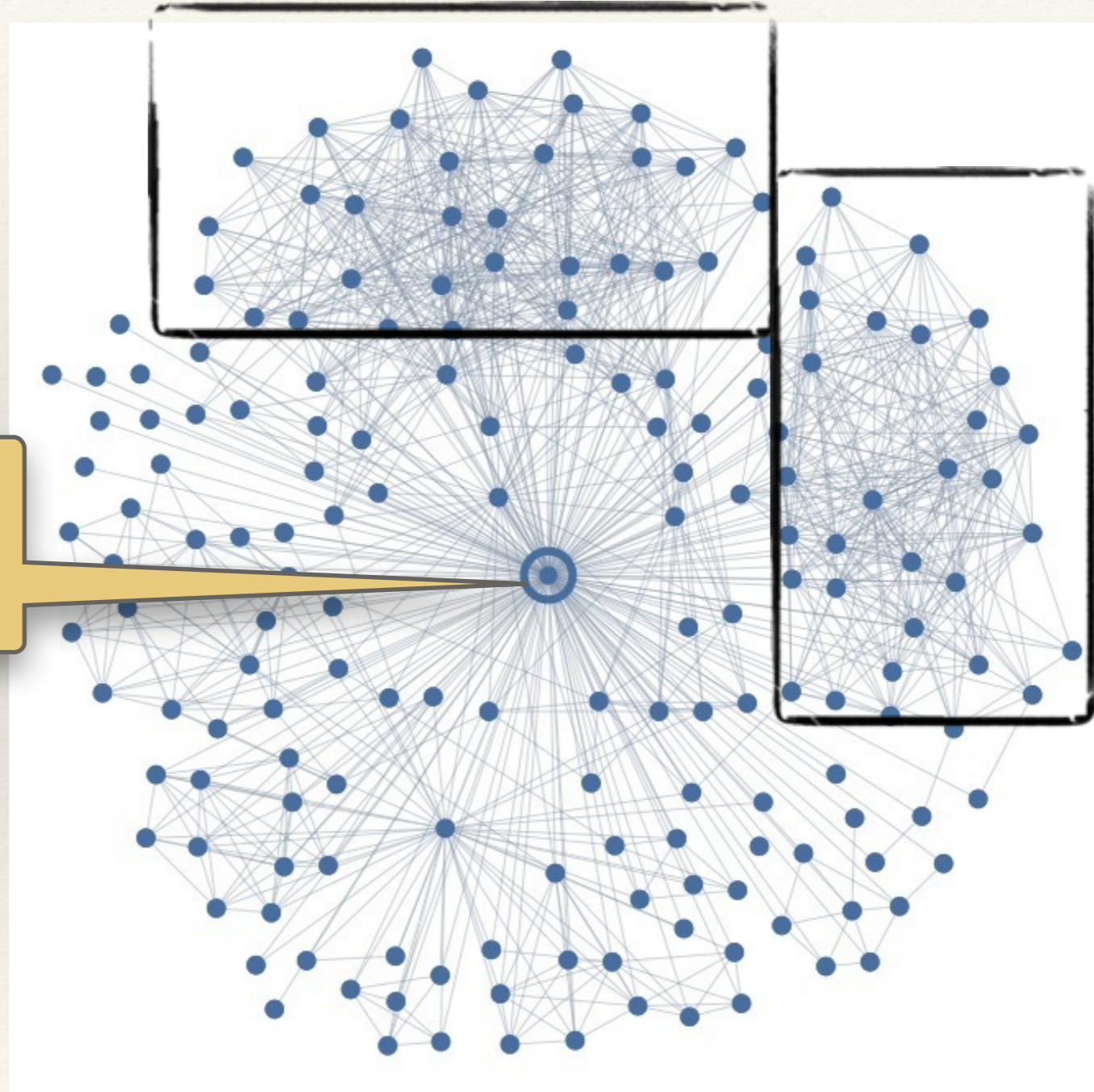
Choose the right tool for the job

- Remember canonical cases where methods succeed or fail

  - A1: PCA vs. CCA vs. RP

  - clustering: k-means likes "round" clusters; Gaussians are good for "ellipses

- Use **wishful thinking** to your advantage to create new tools!

http://www.alice-in-wonderland.net/pictures/tweedledee-tweedledum-pictures.html

Ask: do subtle differences matter?

- A1: rotations
- Clustering definitions
- A2: additions of a single point
- Maybe you can come up with new techniques…

http://www.alice-in-wonderland.net/pictures/tweedledee-tweedledum-pictures.html
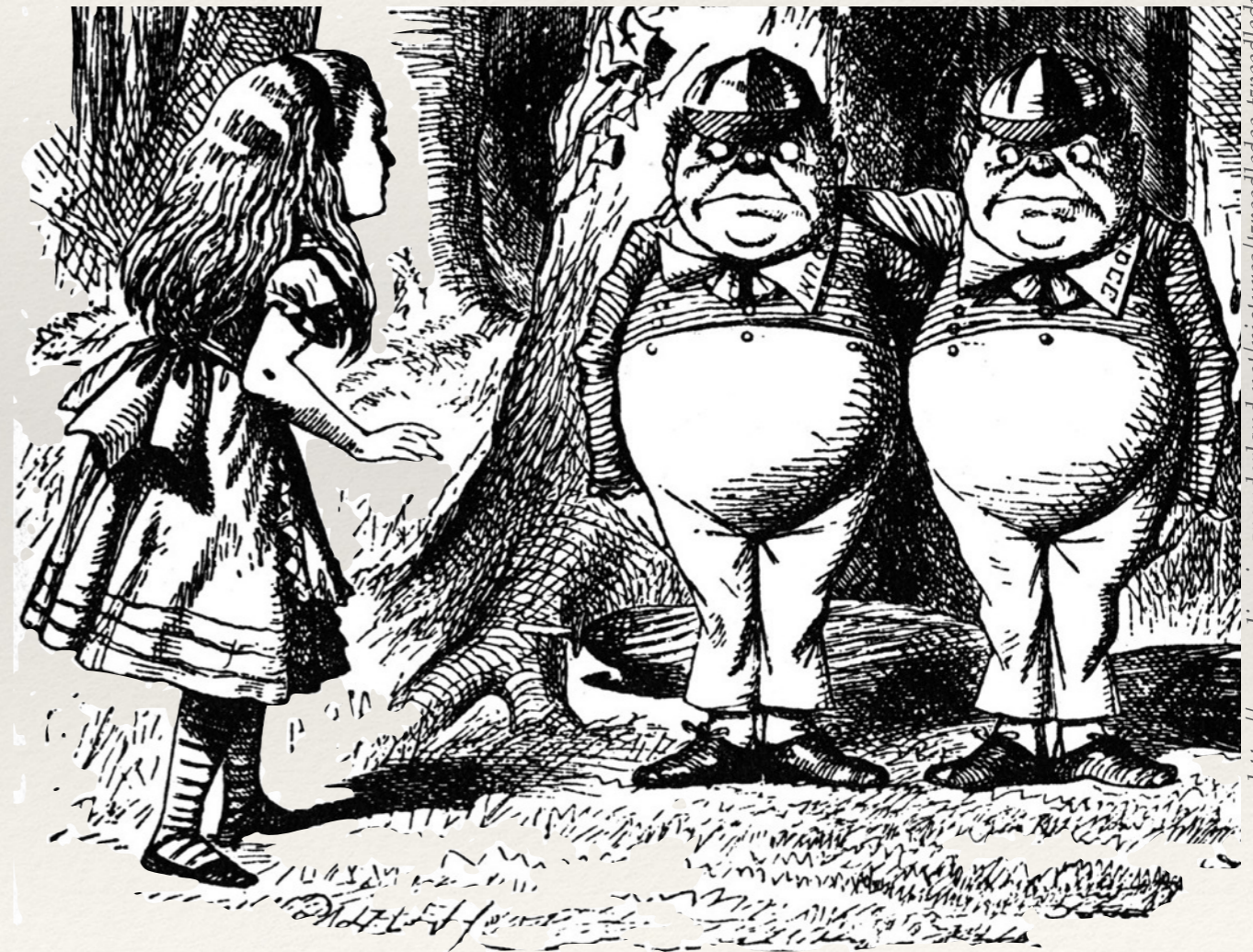
Ask: do subtle differences matter?

the ADFFFTT model in A3

❖ A1: rotations

❖ Clustering definitions

❖ A2: additions of a single point

❖ Maybe you can come up with new techniques…

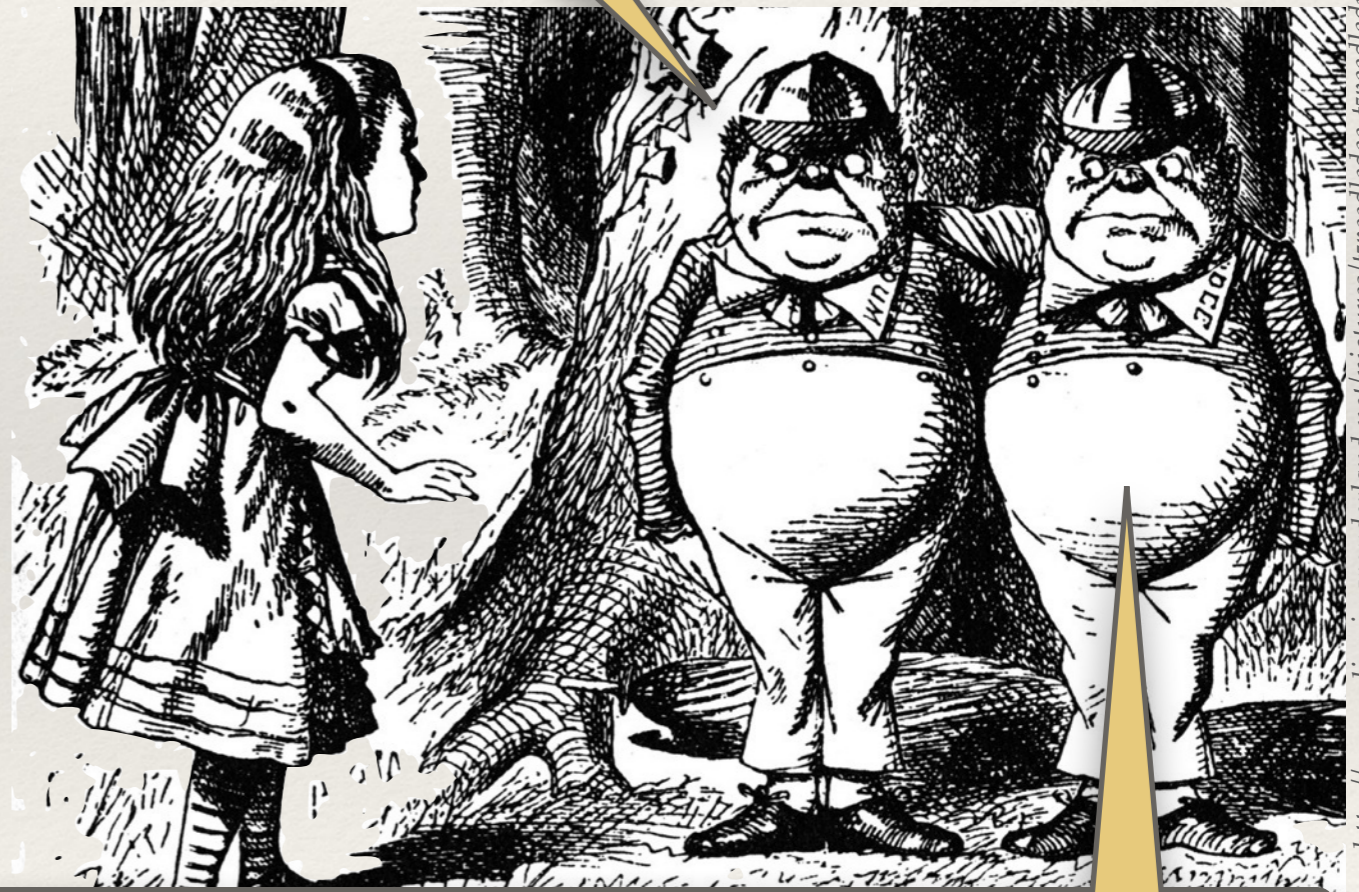# Ask: do subtle differences matter?

the ADFFFTT model in A3

❖ A1: rotations

❖ Clustering definitions

❖ A2: additions of a single point

❖ Maybe you can come up with new techniques…

the *other* ADFFFTT model in A3

"Begin at the beginning," the King said, very gravely, "and go on till you come to the end: then stop."

http://www.cleavebooks.co.uk/grol/alice/won00.htm

"Begin at the beginning," the King said, very gravely, "and go on till you come to the end: then stop."

13

Prof Lee, with lecture notes

Prof Sridharan

"Begin at the beginning," the King said, very gravely, "and go on till you come to the end: then stop."

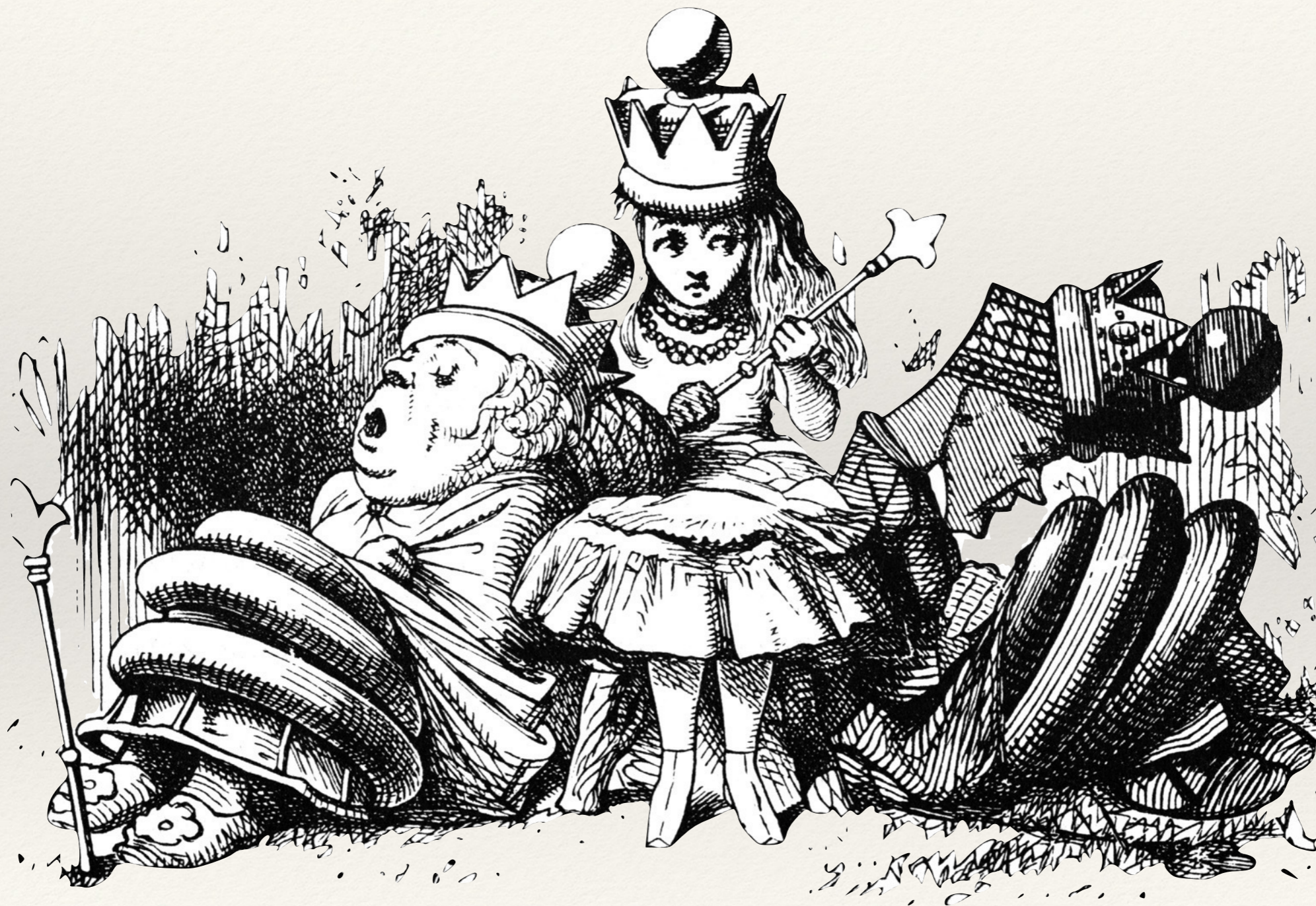http://www.cleavebooks.co.uk/grol/alice/won00.htm

13

Prof Lee, with lecture notes

Prof Sridharan

you guys

"Begin at the beginning," the King said, very gravely, "and go on till you come to the end: then stop."

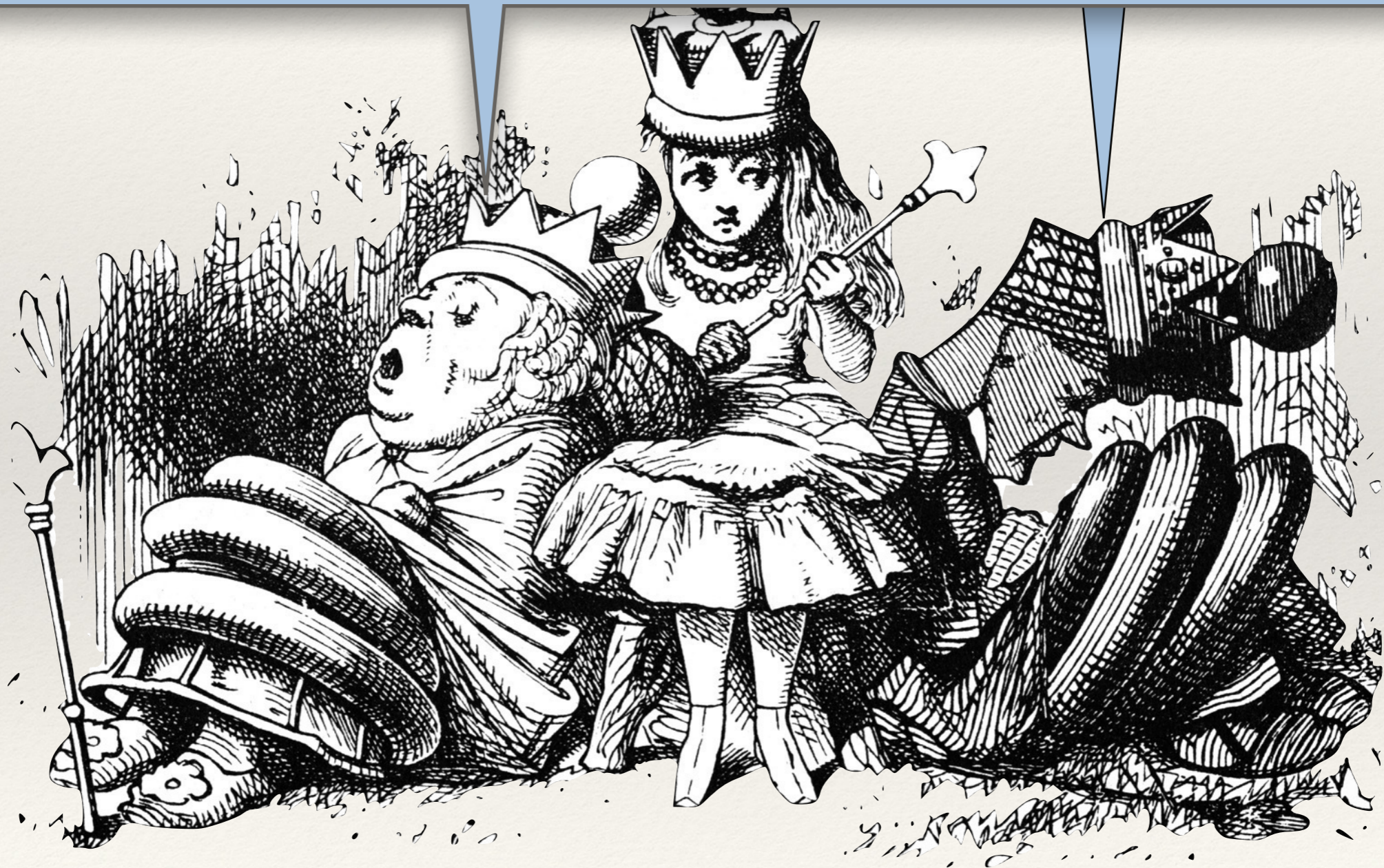http://www.cleavebooks.co.uk/grol/alice/won00.htm

13

You've had to put up with a lot this semester

You've had to put up with a lot this semester

We've been up all night preparing the next lecture/assignment/…

14

You've had to put up with a lot this semester

We've been up all night preparing the next lecture/assignment/…

Man, bringing this iClicker to class is getting annoying.

http://www.……nd.net/alice2b.html

Thanks for …

- …coming to the weekend lecture with Lars Backstrom
- … helping each other on Piazza
- …your patience with us during our first running of this course

Thanks for ...

- ...coming to the weekend lecture with Lars Backstrom
- ... helping each other on Piazza
- ...your patience with us during our first running of this course

You guys rule!

http://www.alice-in-wonderland.net/alice2b.html