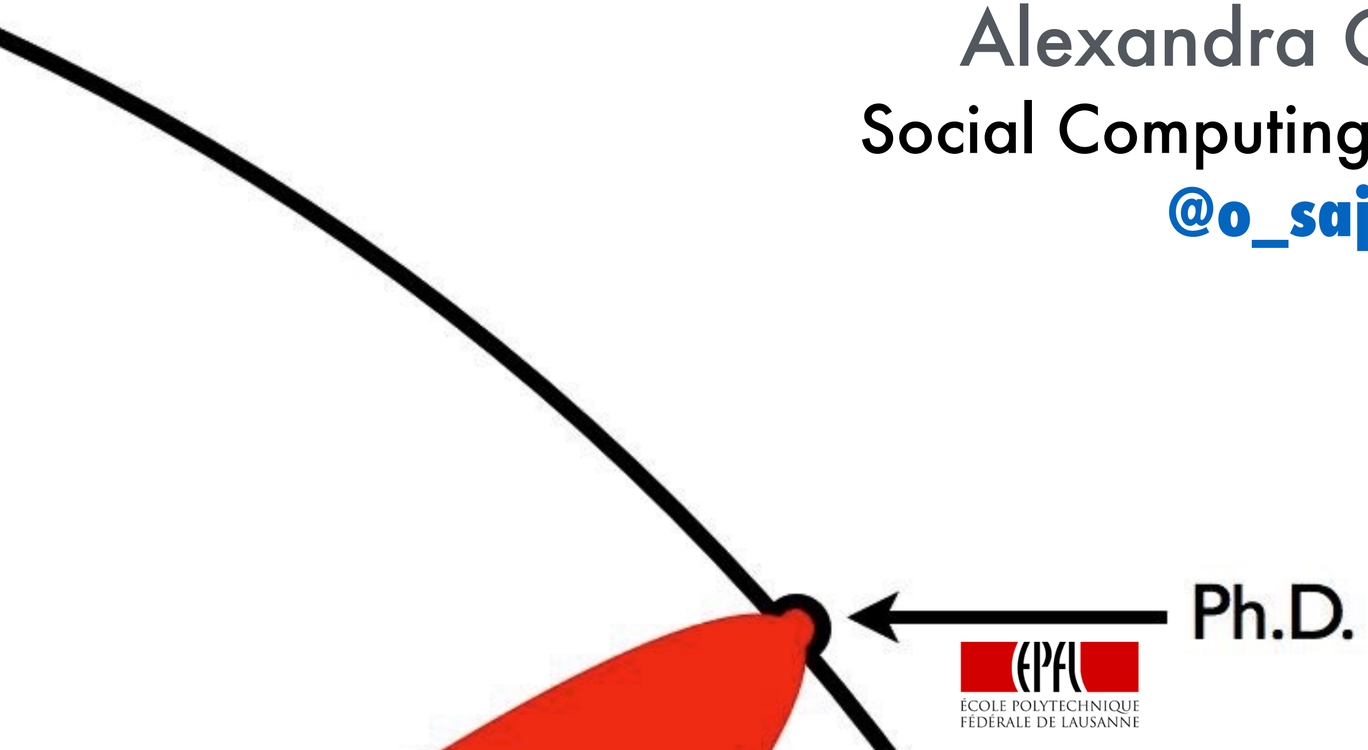
Challenges, Problem Selection, and Impact:
A Biased Perspective on PhD Studies

Alexandra Olteanu
Social Computing Researcher
@o_saja

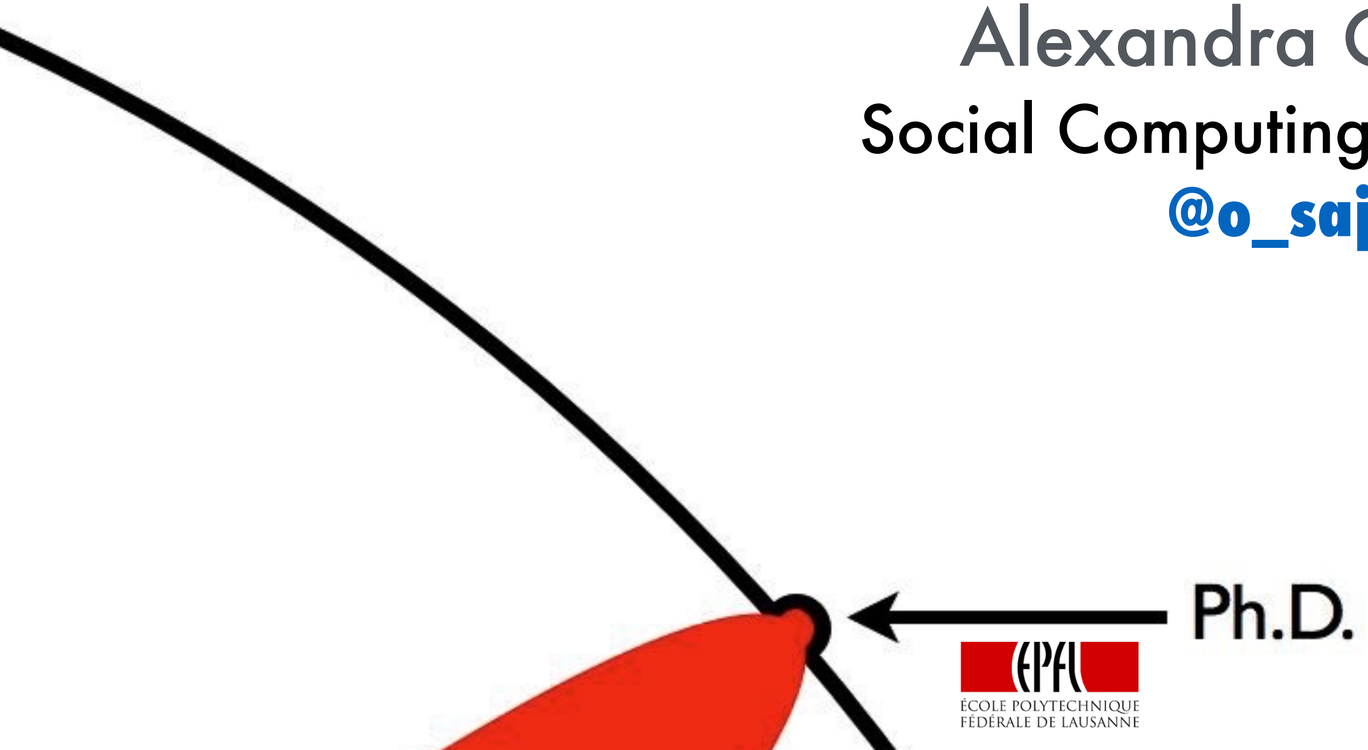


ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

Challenges, Problem Selection, and Impact: A Biased Perspective on PhD Studies

Alexandra Olteanu
Social Computing Researcher
@o_saja



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

The Illustrated Guide to Ph.D. by Matt Might

All it matters is to keep pushing!

Source: http://matt.might.net/articles/phd-school-in-pictures/IllustratedGuidePhD-Matt-Might.pdf

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

My Topical Evolution

B.Sc. Engineering/Specialization in System Programming

kernel programming, compilers, distributed programming, etc.

2 x **M.Sc.** Parallel and Distributed Computer Systems

P2P Social Networks are feasible from an architectural standpoint

Ph.D. it's complicated

started working on web credibility

slowly shifted towards social networks and recommender systems

now interested in using computational tools for social good

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

Quantifying My Topical Evolution

- Started my Ph.D. in September 2011
- Raw counts:
  - 7 full papers (+ 1 forthcoming survey paper)
    - CSCW'17/'15, ICWSM'15/'14, WISE'14, ECIR'13, CIKM'12
  - 6 Others (demo, workshop, symposia, extended abstract, tutorial, etc.)
    - ICWSM'16, AAAI Spring Symposia'16, CHI'13, ICDM'12, EuroSys'12
  - 2 Best Paper Awards
  - 1st try acceptance rate: 85% (90% before my private defense)
  - 18 co-authors
- CrisisLex.org, released over 30 data collections

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

Quantifying My Topical Evolution

- Started my Ph.D. in September 2011
- Raw counts:
  - 7 full papers (+ 1 forthcoming survey paper)
    - CSCW'17/'15, ICWSM'15/'14, WISE'14, ECIR'13, CIKM'12
  - 6 Others (demo, workshop, symposia, extended abstract, tutorial, etc.)
    - ICWSM'16, AAAI Spring Symposia'16, CHI'13, ICDM'12, EuroSys'12
  - 2 Best Paper Awards
  - 1st try acceptance rate: 85% (90% before my private defense)
  - 18 co-authors
- CrisisLex.org, released over 30 data collections

ICDM 2016
BARCELONA
IEEE International Conference on Data Mining

What Sort of Talk is This?

Topics

PhD Challenges

Collaborating

Problem Selection

Impact

PhD in Phases

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

What Sort of Talk is This?

Topics

PhD Challenges

Collaborating

Problem Selection

Impact

PhD in Phases

How?

Serious

Representative

Big/Actionable Insights

Data Mining

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

What Sort of Talk is This?

Topics

PhD Challenges

Collaborating

Problem Selection

Impact

PhD in Phases

How?

Don't take yourself too seriously!

Representative

Big/Actionable Insights

Data Mining

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

What Sort of Talk is This?

Topics

PhD Challenges

Collaborating

Problem Selection

Impact

PhD in Phases

How?

Don't take yourself too seriously!

Just my personal, biased perspective

Big/Actionable Insights

Data Mining

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

What Sort of Talk is This?

Topics

PhD Challenges

Collaborating

Problem Selection

Impact

PhD in Phases

How?

Don't take yourself too seriously!

Just my personal, biased perspective

~~Big~~/Actionable Insights

Data Mining

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

What Sort of Talk is This?

Topics

PhD Challenges

Collaborating

Problem Selection

Impact

PhD in Phases

How?

Don't take yourself too seriously!

Just my personal, biased perspective

~~Big~~/Actionable Insights

Applications of Data Mining

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

Challenges (& Opportunities) in …

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

…the Life of a PhD Student

Deadlines

Writing

Talks

Grading

TODOs

…the Life of a woman PhD Student

Deadlines

Writing

Talks

Grading

TODOs

Story of My (PhD) Life

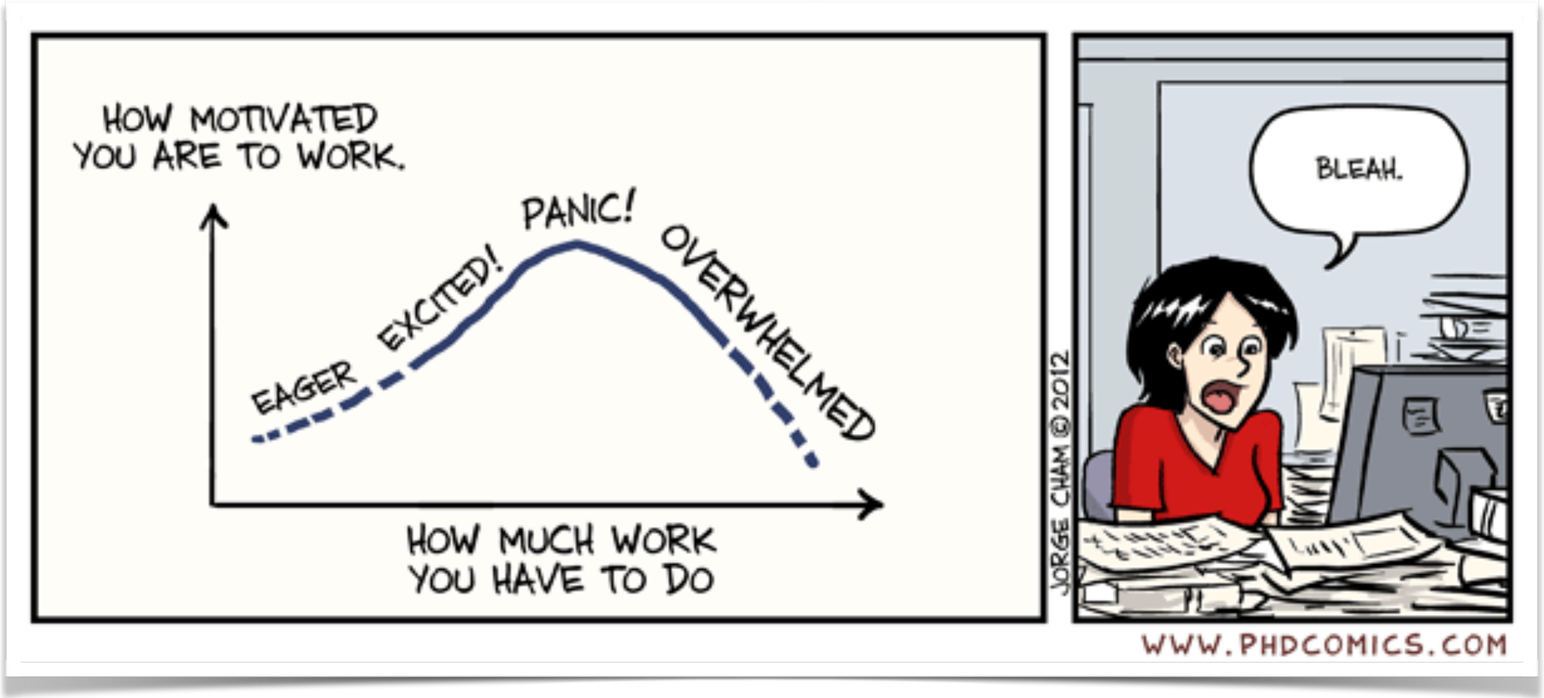
I don’t like deadlines :’-(

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

Story of My (PhD) Life

I don't like deadlines :'-(

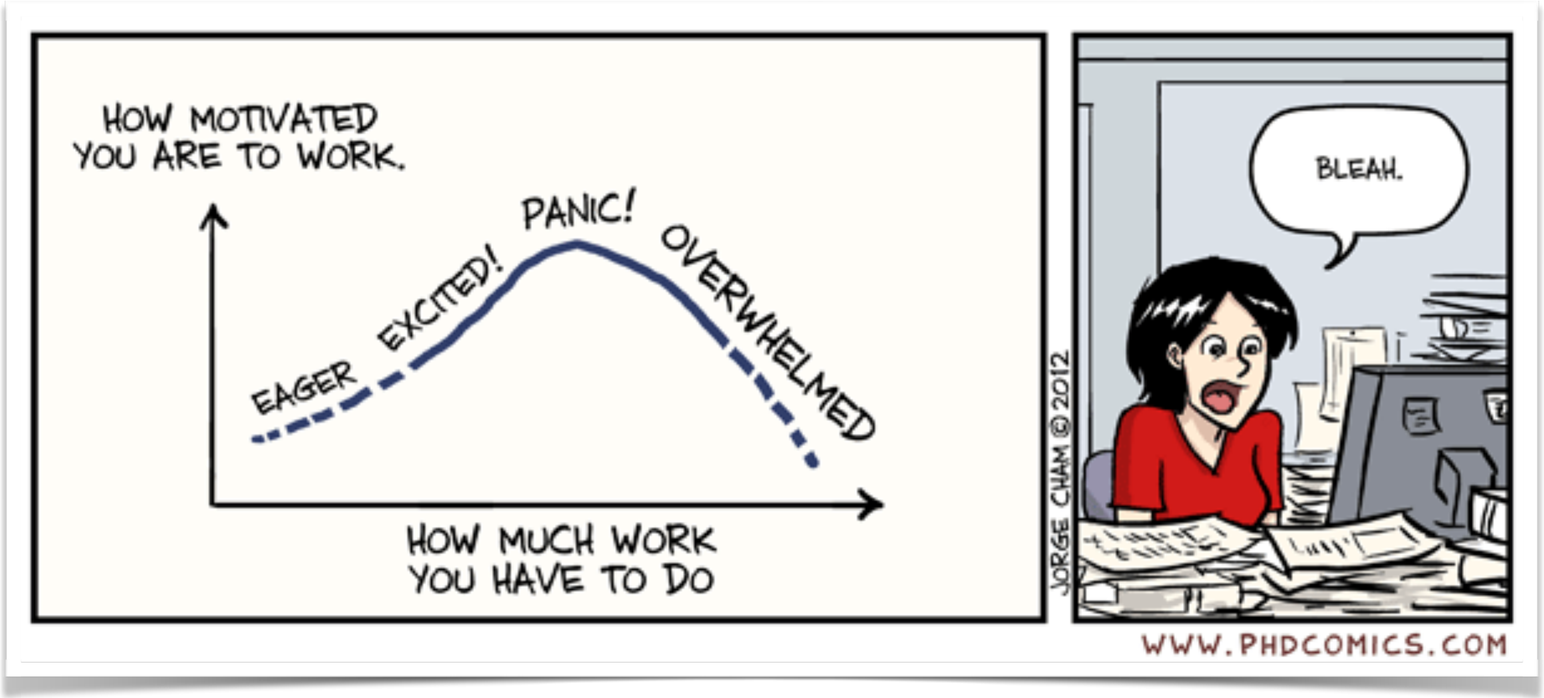
...but I would have been so much less productive without them ...



Story of My (PhD) Life

I don’t like deadlines :’-(

…but I would have been so much less productive without them …



Use them wisely!

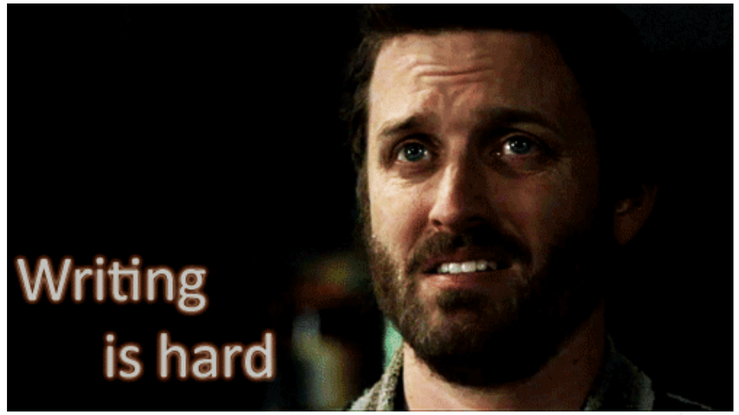
ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

Writing Is Hard, But …

You are responsible for properly communicating your work.

Copy editing is not fun, so **be kind with your advisor** and senior colleagues.

It gets better in time: **the more you write, the better you write**.



Dry-runs, Damned Dry-runs & Speaker Notes

- Rely on well-rehearsed spontaneity,
- or on speaker-notes based spontaneity
- Synchronized animations are better than the laser pointer
- My speaker notes contain <click> marks for animation builds

Dry-runs, Damned Dry-runs & Speaker Notes

- Rely on well-rehearsed spontaneity,
- or on speaker-notes based spontaneity
- Synchronized animations are better than the laser pointer
- My speaker notes contain <click> marks for animation builds

Dry-runs, Damned Dry-runs & Speaker Notes

- Rely on well-rehearsed spontaneity,
- or on speaker-notes based spontaneity
- Synchronized animations are better than the laser pointer
- My speaker notes contain <click> marks for animation builds

Put <u>mathematical formulas</u> on slides iff …

- you spend over 30 seconds for each variable, and
- your audience have seen them before, and
- your are an *one-(wo)man-show*

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

Dry-runs, Damned Dry-runs & Speaker Notes

- Rely on well-rehearsed spontaneity,
- or on speaker-notes based spontaneity
- Synchronized animations are better than the laser pointer
- My speaker notes contain <click> marks for animation builds

Put mathematical formulas on slides iff …

- yo… …e, and
- …efore, and
- …your are an one-(wo)*man-show*

Most people just can't parse them in too little time.

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

Dry-runs, Damned Dry-runs & Speaker Notes

- Rely on well-rehearsed spontaneity,
- or on speaker-notes ba[illegible]
- Sync[illegible]ations are better than the laser pointer
- My speaker notes contain <click> marks for animation builds

Optimize for clarity, not precision.

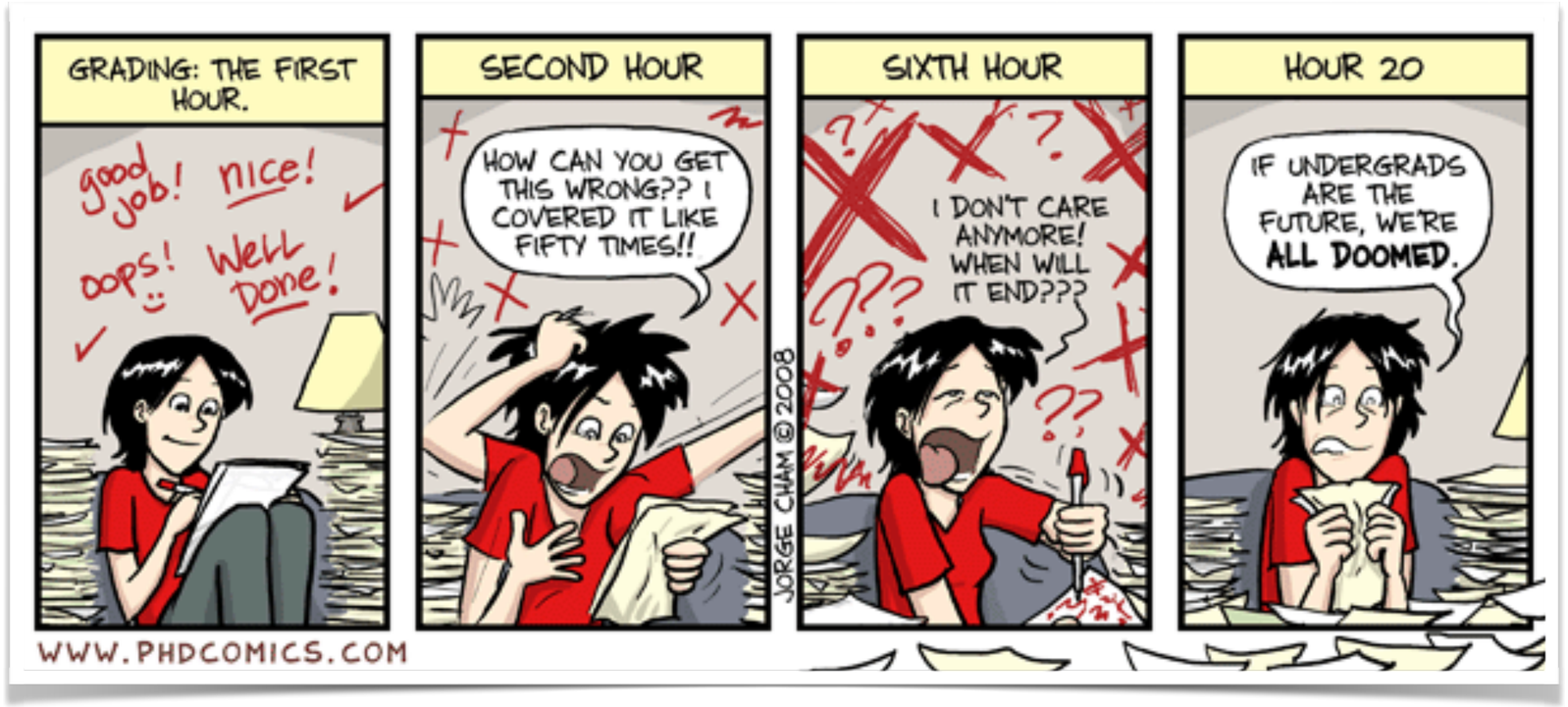
Put <u>mathematical formulas</u> on slides iff …

- yo[illegible]e, and
- [illegible]before, and
- [illegible]your are an one-(wo)man-show

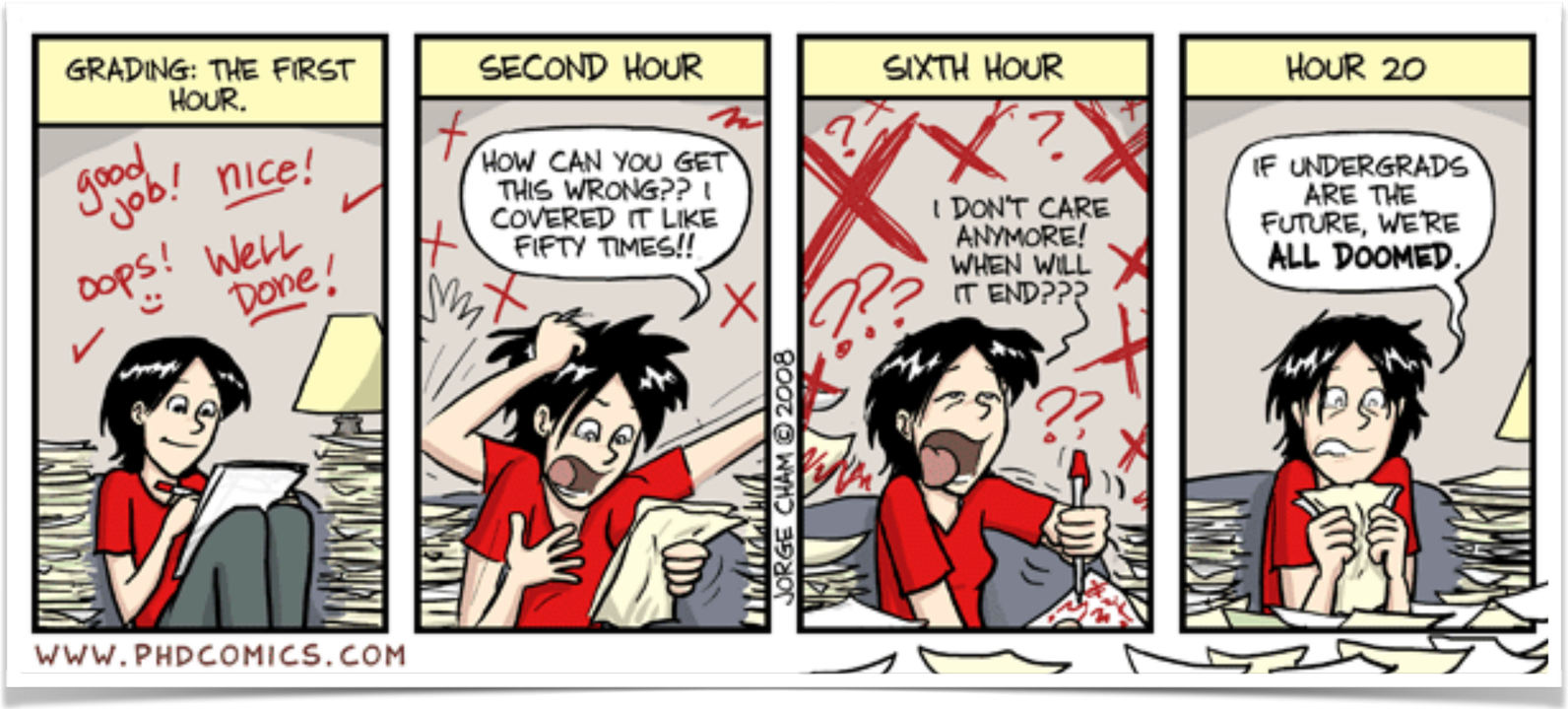
Most people just can't parse them in too little time.

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

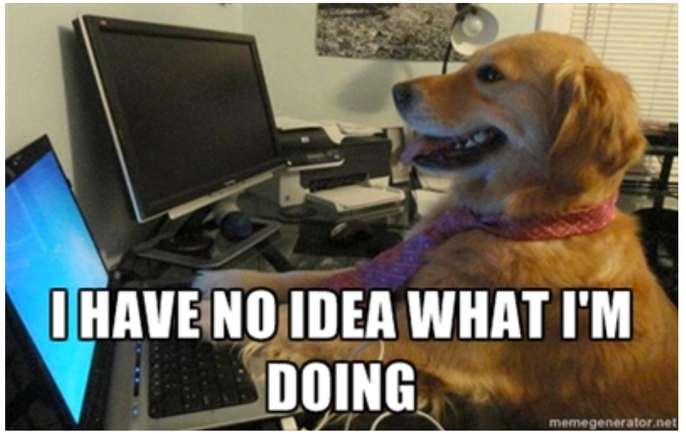
Ohh, the Joy of Grading



Ohh, the Joy of Grading



...but helpful whenever
you feel like an impostor



When You Have Too Many Things TO DO

Worry no more.

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

When You Have Too Many Things TO DO

Worry no more.

Make a TODO list!

ICDM 2016
BARCELONA
IEEE International Conference on Data Mining

Collaboration Is Key.

Collaborators

- Advisor(s)
- Outside mentors
  - from internships
  - shared interests
  - visiting professors
- Lab mates
- Students

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

Collaborators

- Advisor(s)
- Outside mentors
  - from internships
  - shared interests
  - visiting professors
- Lab mates
- Students

Look for diversity in backgrounds, and respect their perspective.

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

Finding Mentors

- Remember
  - they are busy
  - they have their own priorities/interests
- Align your goals with their interests
- Do not overestimate their benefits
- They can help with more than research
  - networking, support, endorsement, etc.

Finding Mentors

- Remember
  - they are busy
  - they have their own priorities/interests
- Align your goals with their interests
- Do not overestimate their bene
- They can help with mo rch
  - networking, support , etc.

But, keep them accountable for the things they commit to.

Be a Mentor

- Take their interests into account
- Advisor shoes
- Put them first
  - review their work ⇒ will help you write better.
  - do dry-runs with them ⇒ will help you make better presentations.
- You learn better when you explain it to others

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

Be a Mentor

- Take their interests into account
- Advisor shoes
- Put them first
  - review their work ⇒ will help you write better.
  - do dry-runs with them ⇒ will help you
    presentations.
- You learn better when y
  ners

Offer help, do not expect them to always know what they need.

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

Receiving Feedback: A Few "Do not(s)"

- Don't be critical of how people give you (solicited) feedback
  - "You do not know how to give feedback"
- Don't get defensive and argumentative
  - "You are wrong"
  - "You did not understand it"
- Don't try to figure it out on the spot, take time

Receiving Feedback: A Few "Do not(s)"

- Don't be critical of how people give you (solicited) feedback
  - "You do not know how to give feedback"
- Don't get defensive and argumentative
  - "You are wrong"
  - "You did not understand it"
- Don't try to figure it out on the spot, take time

Be grateful!

Give Credit, Always

Err on the side of giving others more credit, than less credit

Ignoring it, likely the easiest way to make people resentful

It's easy to think your contributions are more valuable than of others

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

Problem Selection

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

Types of Contributions

- A more efficient solution to a known problem
- An interesting solution to a known problem
- Introduces a new problem
  - the solution matters less
- Makes the community contributions more accessible
- Clarifies the trade-offs and gaps of existing approaches
  - includes negative results, meta-analyses, etc.

Some of My Rationales

- Impact more important than being faithful to a topic
  - I decided the topic of my thesis with 3 months before defense
- Do what others don't want to do
- Optimize for relevance within the application domain
- Focus on what can be done, not on what cannot be done
- Illuminating a problem as important as solving a problem

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

Some of My Rationales

- Impact more important than being faithful to a topic
  - I decided the topic of my thesis with 3 months before defense
- Do what others don't want to do
- Optimize for relevance within the application domain
- Focus on what can be done, not on what cannot be done
- Illuminating a problem as important as solving a problem

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

Impact Can Have Many Flavours.

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

How Impact Can Look Like?

- Scientific impact/make a breakthrough
- Popularize a methodological approach
- Enable others to do more/better work
  - release data, tools, surveys, etc.
- Great teaching material
- Policy impact/news coverage

What Is Useful For Those Outside DM

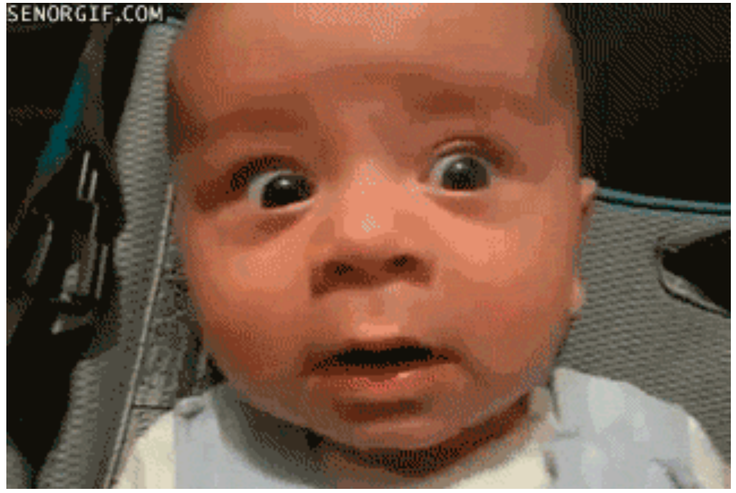
- Do not just release code, tell me when it is (not) useful
  - e.g., textual data vs. images, good for imbalanced/small data
- Give the high-level intuition of your work/solution
  - if it proves to be what I need, I'll check the details
- Go beyond abstract evaluation metrics
  - translate x% improvement in precision in measurable benefits within some application domains

Finally, a Few Ph.D. Phases

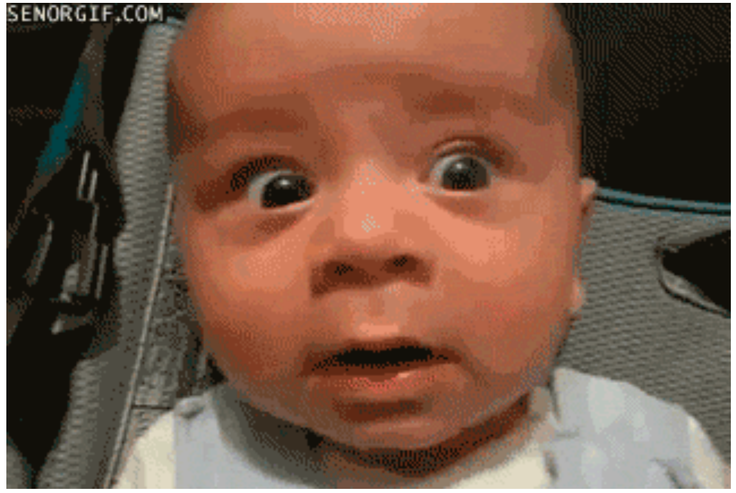
Phase 0: The Before Excitement

Phase 1: The 1st Year Confusion

Phase 1: The 1st Year Confusion



Phase 1: The 1st Year Confusion



But, eventually,
you'll figure it out.

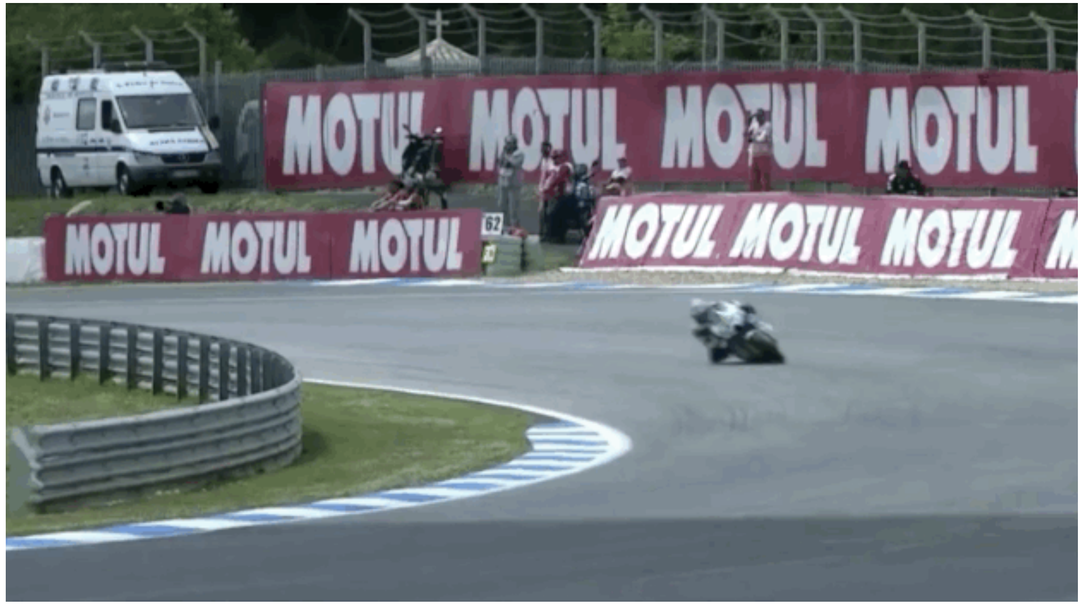
Phase 2: Mid-PhD Crisis

What you say…

and, how you feel…

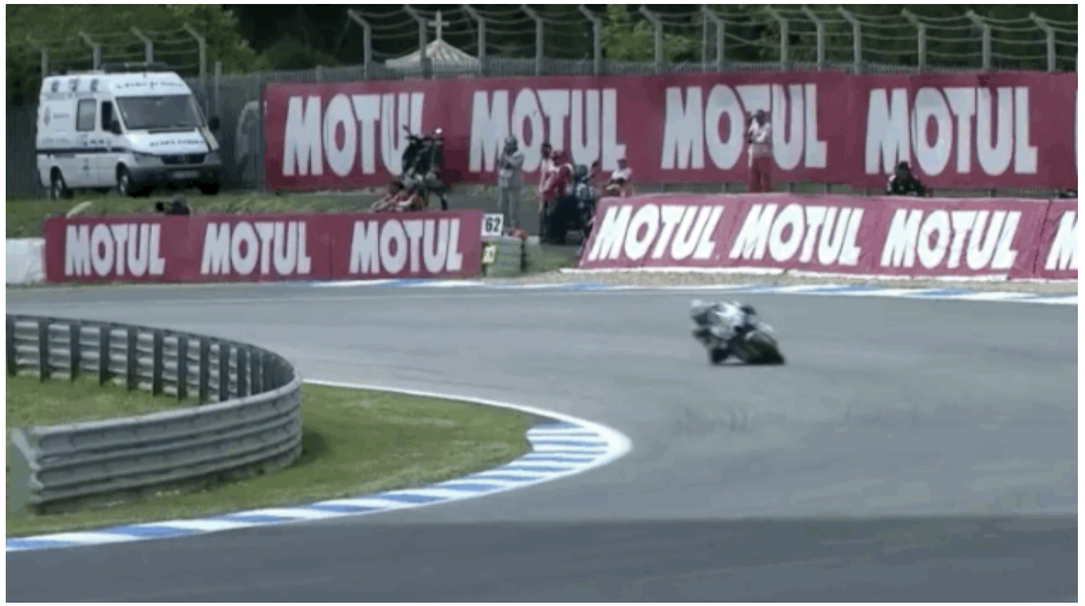
Phase 3: The Serendipity

Often,
the ride is bumpy



Phase 3: The Serendipity

Often,
the ride is bumpy



Sometimes,
you'll feel lucky

Phase 4: Overwhelmed

Sometimes, it feels like
a lot to handle

Phase 4: Overwhelmed

Sometimes, it feels like a lot to handle

But, you'll pull it through

Last Phase: What Is the Next Step?

KEEP CALM AND QUESTION EVERYTHING

Email:
alexandra@aolteanu.com

Web:
aolteanu.com

Twitter:
@o_saja

KEEP
CALM
AND
QUESTION
EVERYTHING