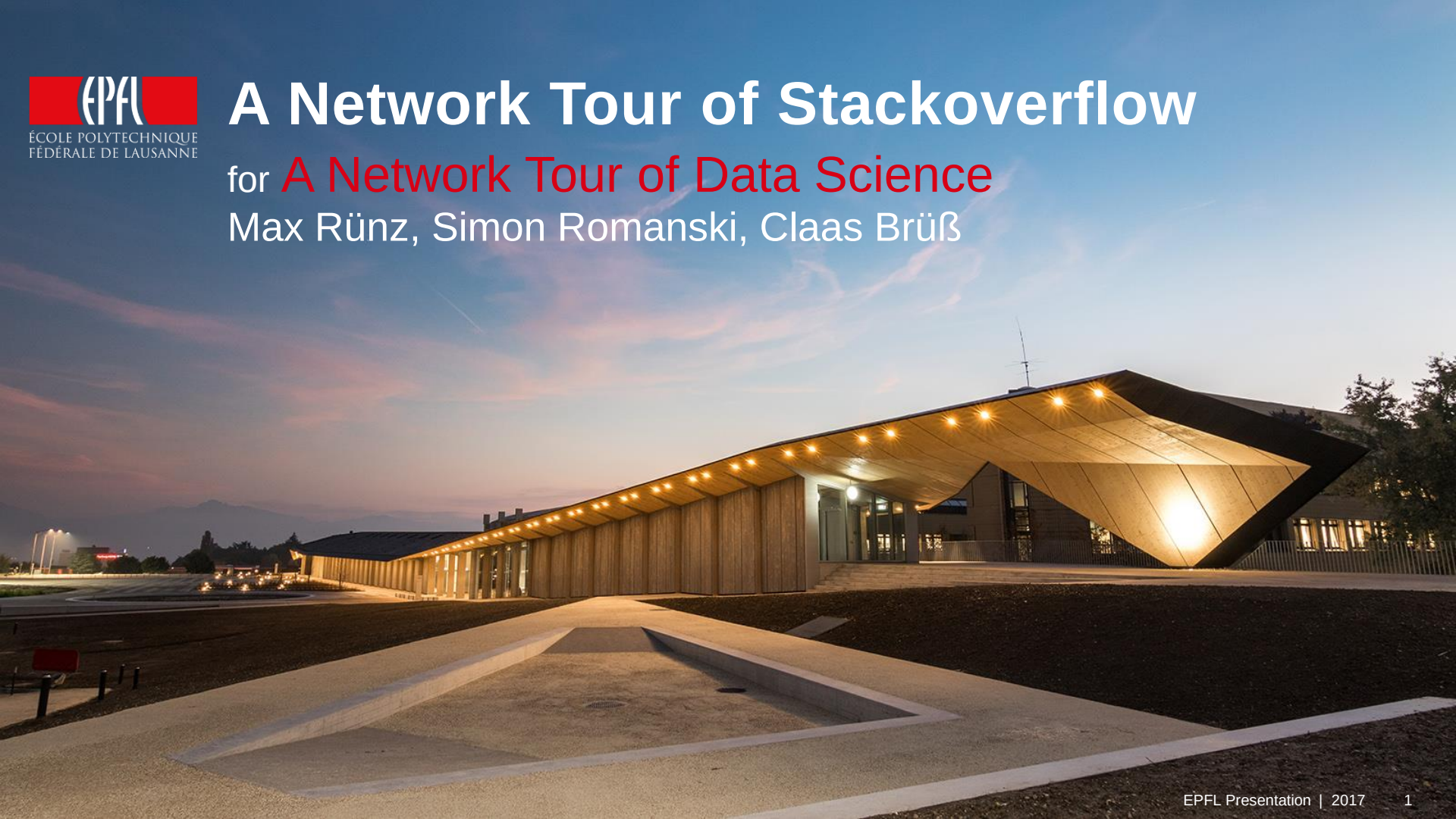
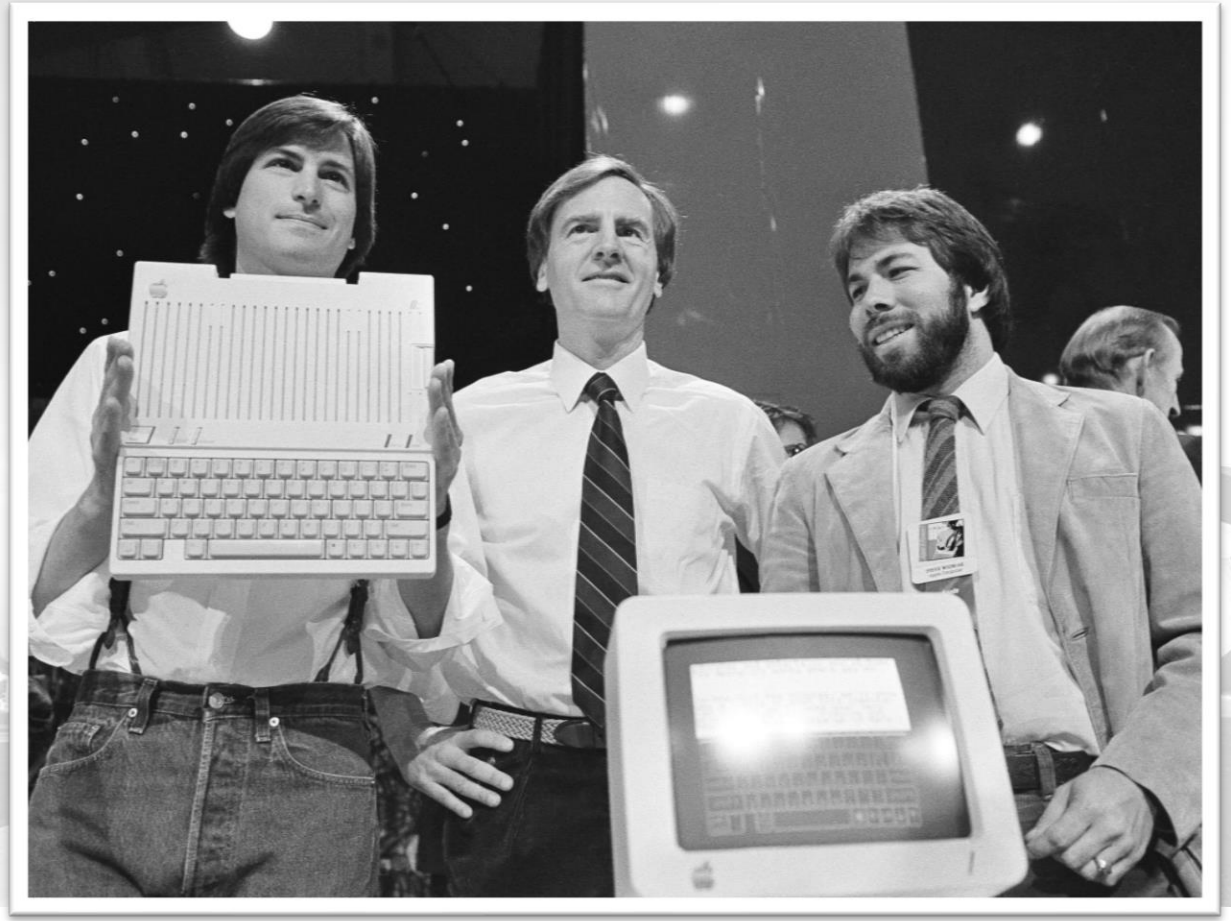
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

A Network Tour of Stackoverflow

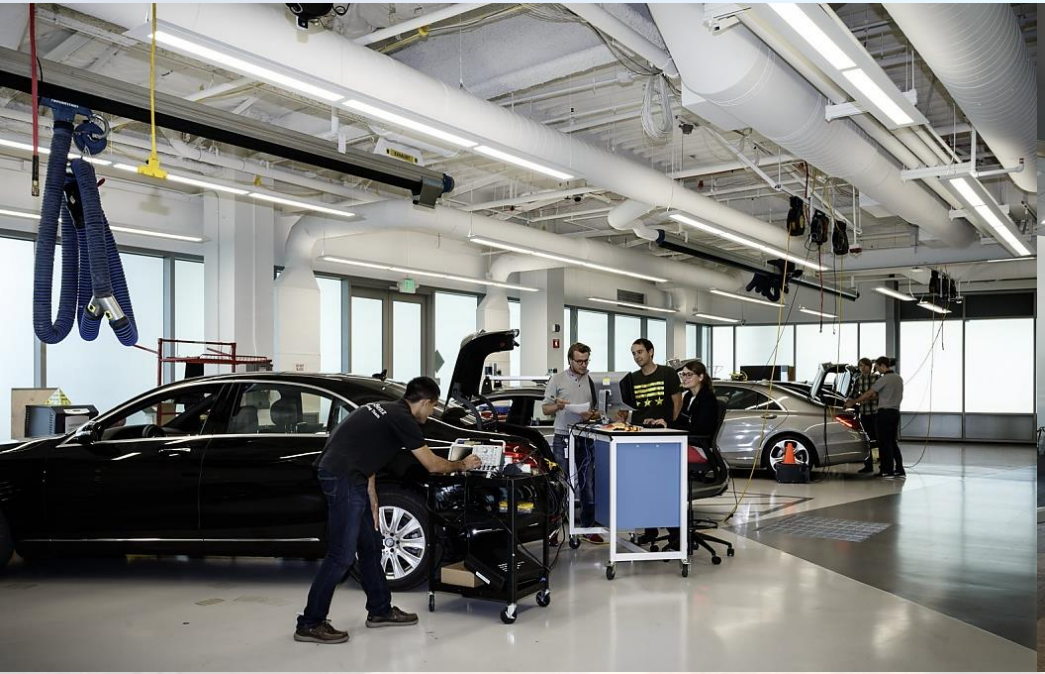
for A Network Tour of Data Science

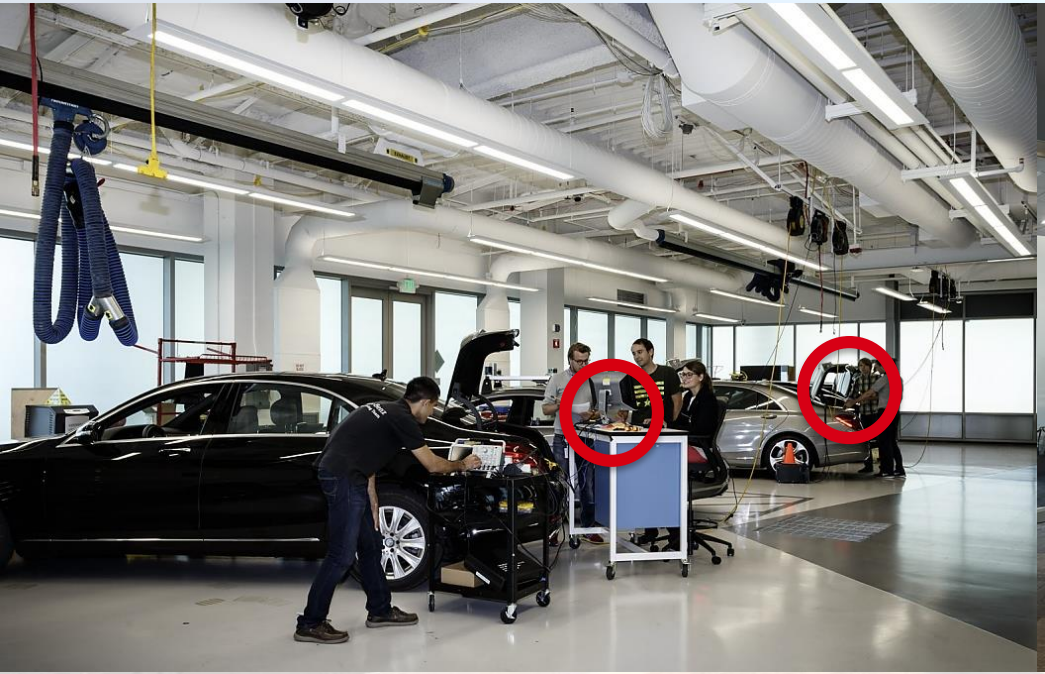
Max Rünz, Simon Romanski, Claas Brüß

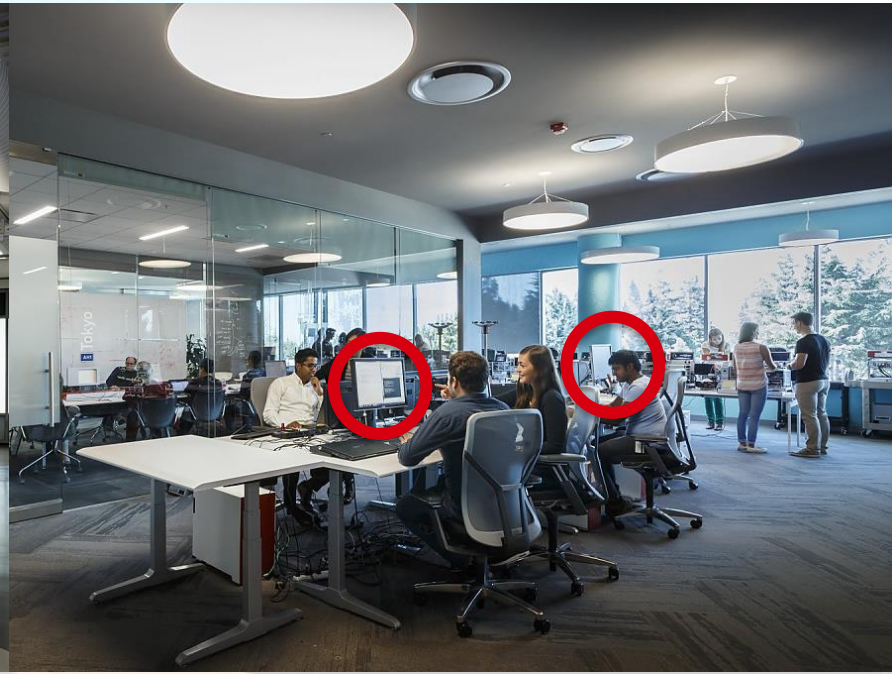


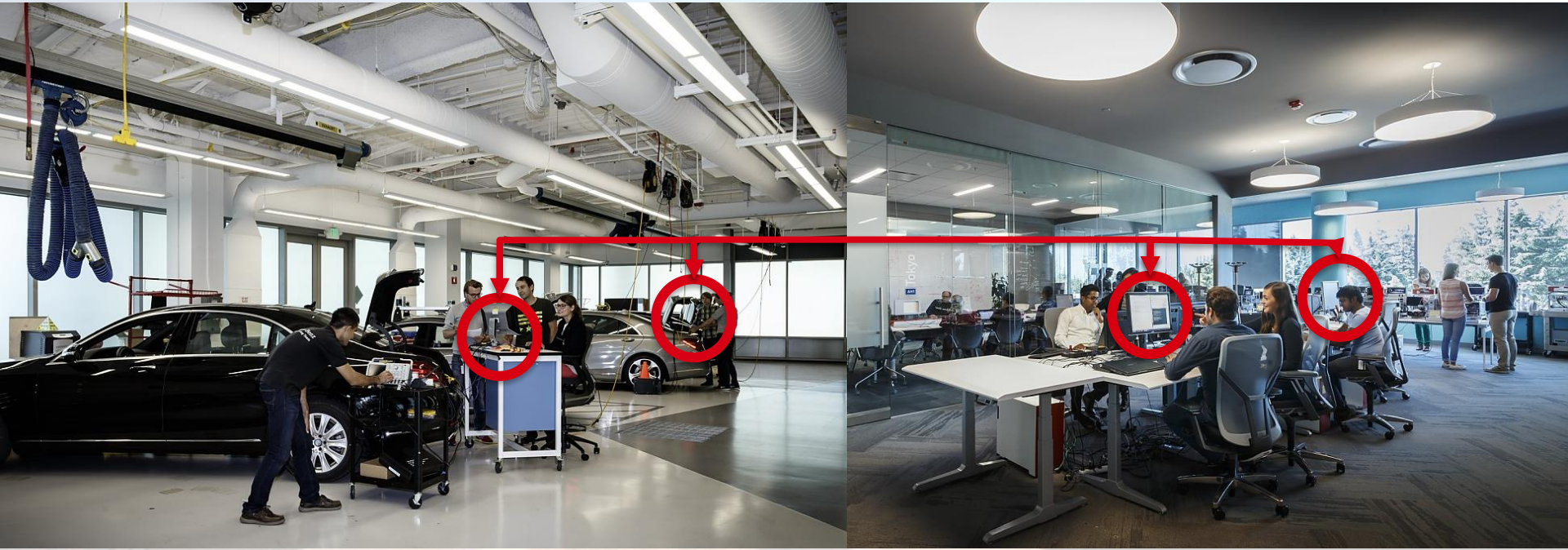



FEBRUARY 20, 2006
WHO'S BEHIND THE CARTOON MAYHEM? BEING OBAMA
TIME
CAN WE TRUST
Google
WITH OUR SECRETS?
An exclusive inside look at the $100 billion empire that is dominating the Internet
BY ADI IGNATIUS
Google honchos, from top, Larry Page, Eric Schmidt and Sergey Brin

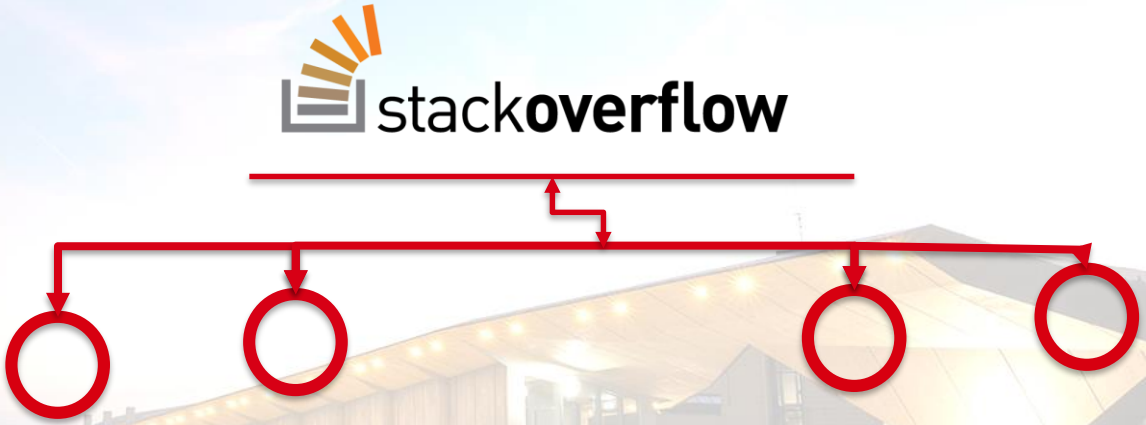




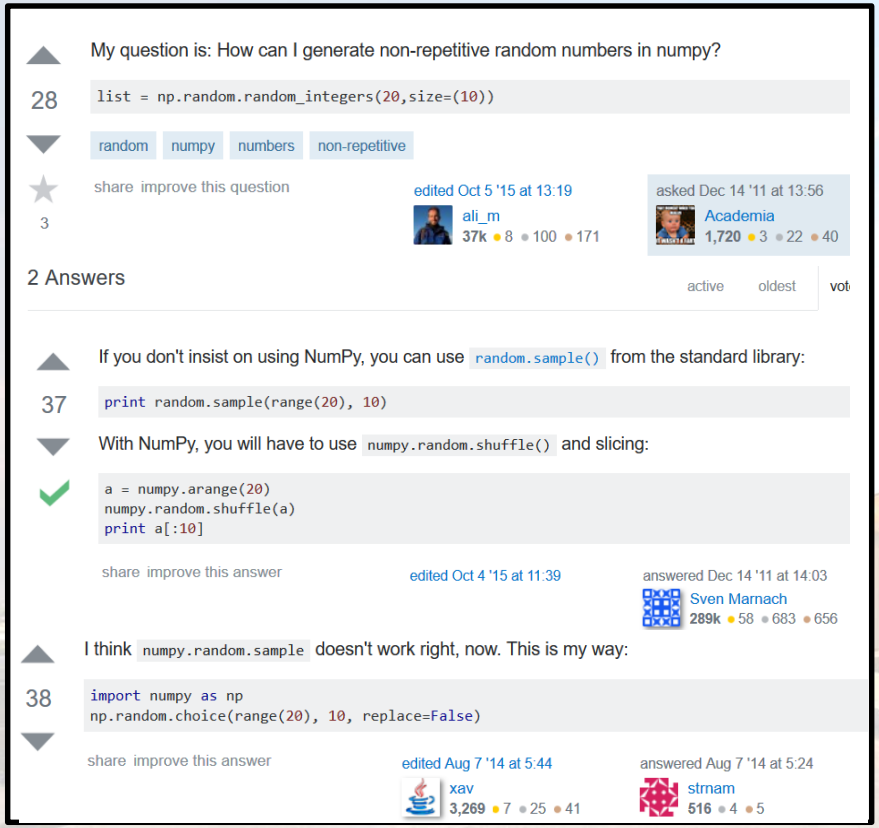




Analysis of the Stackoverflow Collaboration Model



Network Creation



My question is: How can I generate non-repetitive random numbers in numpy?

28

```
list = np.random.random_integers(20,size=(10))
```

random numpy numbers non-repetitive

share improve this question

edited Oct 5 '15 at 13:19
ali_m
37k ● 8 ● 100 ● 171

asked Dec 14 '11 at 13:56
Academia
1,720 ● 3 ● 22 ● 40

3

2 Answers

active oldest vot

If you don't insist on using NumPy, you can use `random.sample()` from the standard library:

37

```
print random.sample(range(20), 10)
```

With NumPy, you will have to use `numpy.random.shuffle()` and slicing:

```
a = numpy.arange(20)
numpy.random.shuffle(a)
print a[:10]
```

share improve this answer

edited Oct 4 '15 at 11:39

answered Dec 14 '11 at 14:03
Sven Marnach
289k ● 58 ● 683 ● 656

I think `numpy.random.sample` doesn't work right, now. This is my way:

38

```
import numpy as np
np.random.choice(range(20), 10, replace=False)
```

share improve this answer

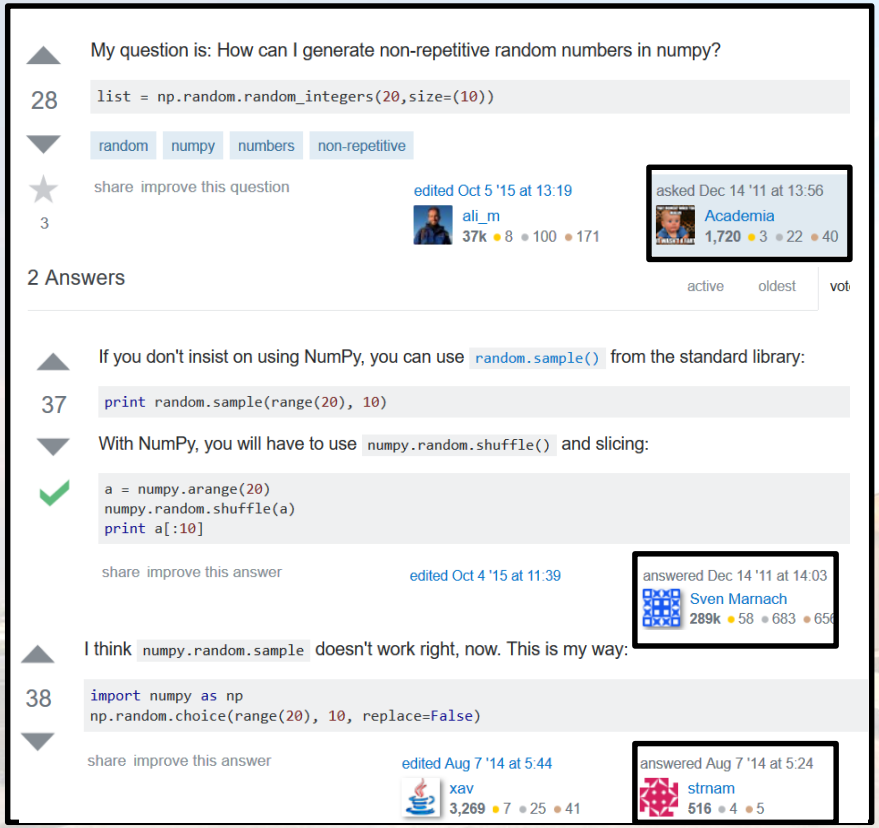
edited Aug 7 '14 at 5:44
xav
3,269 ● 7 ● 25 ● 41

answered Aug 7 '14 at 5:24
strnam
516 ● 4 ● 5

Dataset

- Complete data dump from archive.org
- From 2008 until 2017
- Formatted in XML
- ~ 70 GB
- Posts
  - Text
  - Answers
  - Votes
  - Etc

Network Creation



My question is: How can I generate non-repetitive random numbers in numpy?

28

```
list = np.random.random_integers(20,size=(10))
```

random numpy numbers non-repetitive

share improve this question

edited Oct 5 '15 at 13:19
ali_m
37k ● 8 ● 100 ● 171

asked Dec 14 '11 at 13:56
Academia
1,720 ● 3 ● 22 ● 40

3

2 Answers

active oldest vot

If you don't insist on using NumPy, you can use `random.sample()` from the standard library:

37

```
print random.sample(range(20), 10)
```

With NumPy, you will have to use `numpy.random.shuffle()` and slicing:

```
a = numpy.arange(20)
numpy.random.shuffle(a)
print a[:10]
```

share improve this answer

edited Oct 4 '15 at 11:39

answered Dec 14 '11 at 14:03
Sven Marnach
289k ● 58 ● 683 ● 65

I think `numpy.random.sample` doesn't work right, now. This is my way:

38

```
import numpy as np
np.random.choice(range(20), 10, replace=False)
```

share improve this answer

edited Aug 7 '14 at 5:44
xav
3,269 ● 7 ● 25 ● 41

answered Aug 7 '14 at 5:24
strnam
516 ● 4 ● 5

Network Creation

My question is: How can I generate non-repetitive random numbers in numpy?

28

```
list = np.random.random_integers(20,size=(10))
```

random numpy numbers non-repetitive

share improve this question

edited Oct 5 '15 at 13:19
ali_m
37k ● 8 ● 100 ● 171

asked Dec 14 '11 at 13:56
Academia
1,720 ● 3 ● 22 ● 40

3

2 Answers

active oldest vot

37

If you don't insist on using NumPy, you can use `random.sample()` from the standard library:

```
print random.sample(range(20), 10)
```

With NumPy, you will have to use `numpy.random.shuffle()` and slicing:

```
a = numpy.arange(20)
numpy.random.shuffle(a)
print a[:10]
```

share improve this answer

edited Oct 4 '15 at 11:39

answered Dec 14 '11 at 14:03
Sven Marnach
289k ● 58 ● 683 ● 656

38

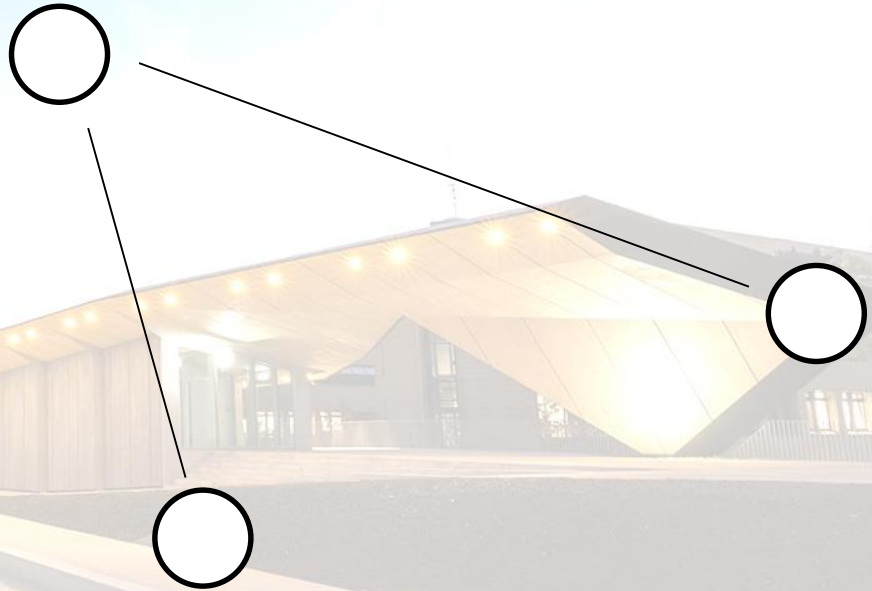
I think `numpy.random.sample` doesn't work right, now. This is my way:

```
import numpy as np
np.random.choice(range(20), 10, replace=False)
```

share improve this answer

edited Aug 7 '14 at 5:44
xav
3,269 ● 7 ● 25 ● 41

answered Aug 7 '14 at 5:24
strnam
516 ● 4 ● 5



Network Creation

Time: 14th Dec 2011
Question Votes: 28
Answer Votes: 37
Accepted: True
Tags: Numpy, Random, …

Time: 7th Aug 2014
Question Votes: 28
Answer Votes: 37
Accepted: True
Tags: Numpy, Random, …

Network Creation

Time: 14th Dec 2011
Question Votes: 28
Answer Votes: 37
Accepted: True
Tags: Numpy

Time: 7th Aug 2014
Question Votes: 28
Answer Votes: 37
Accepted: True
Tags: Numpy

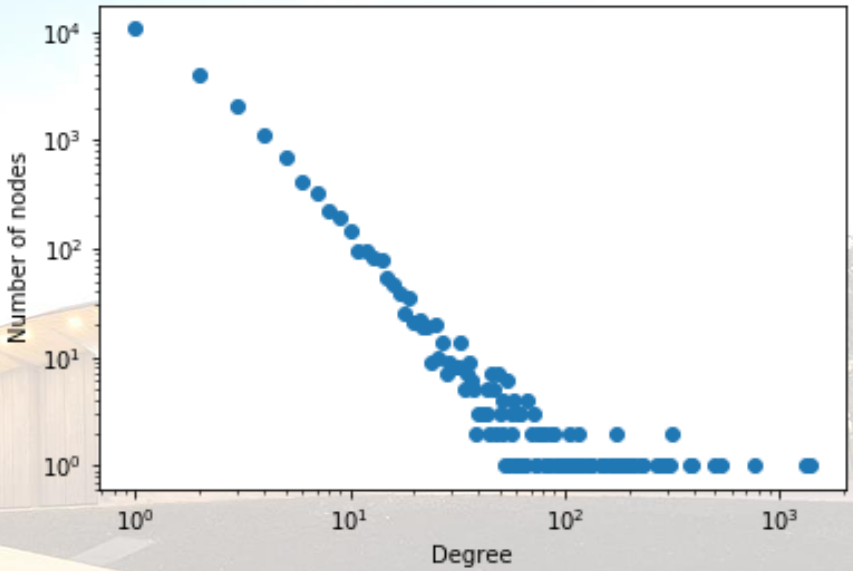
Network Cleaning

- Applied filters
  - Questions with negative votes
  - Answers with negative votes
  - Remove self loops

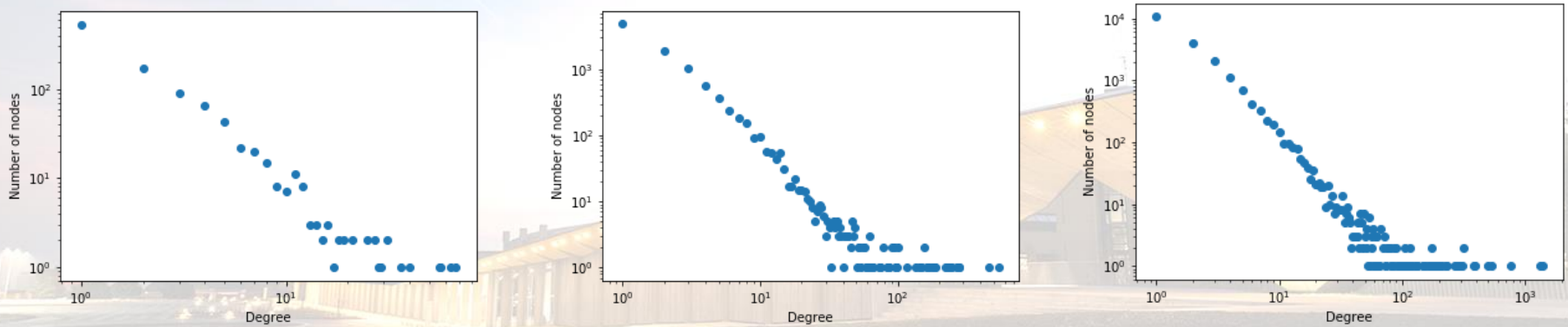
- Considered filters
  - Take giant component
  - Filter by node degree
  - Filter by attributes

Network Exploration

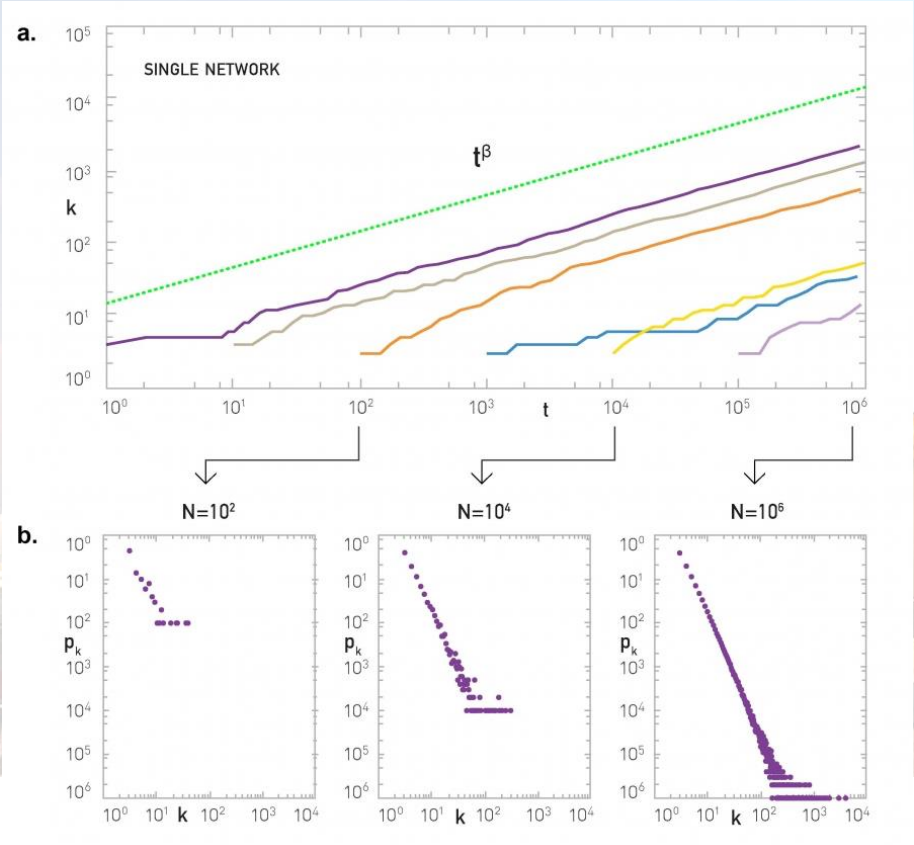
Network Attributes	Value
Nodes	20880
Edges	37231
Conn. Components	999
Giant Component	18934
<k>	3.6618
k_max	1384



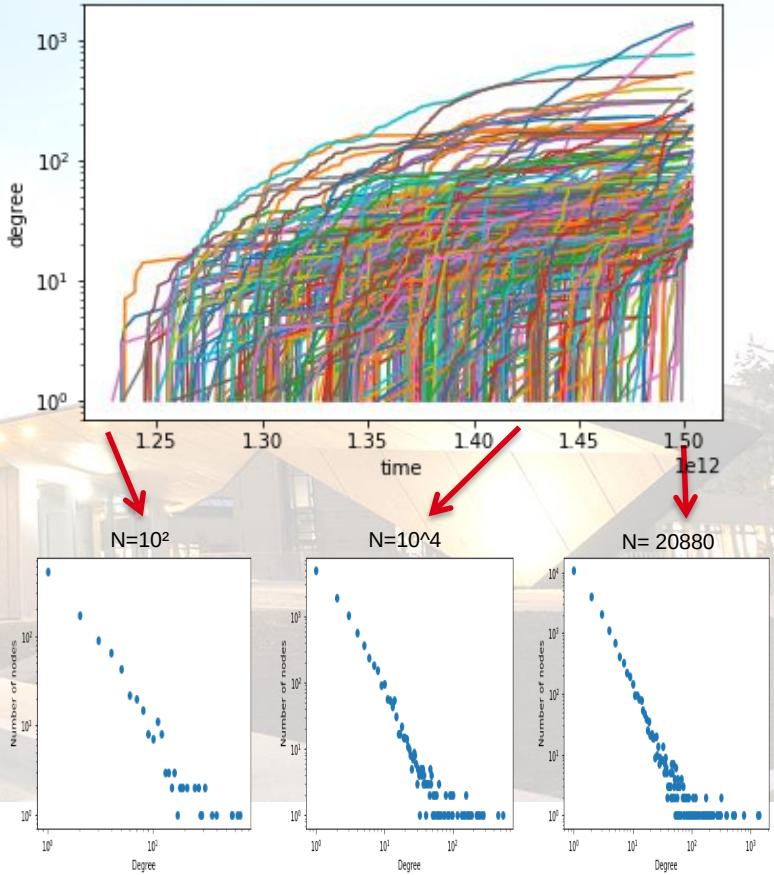
Network Evolution – What kind of Network are we looking at?



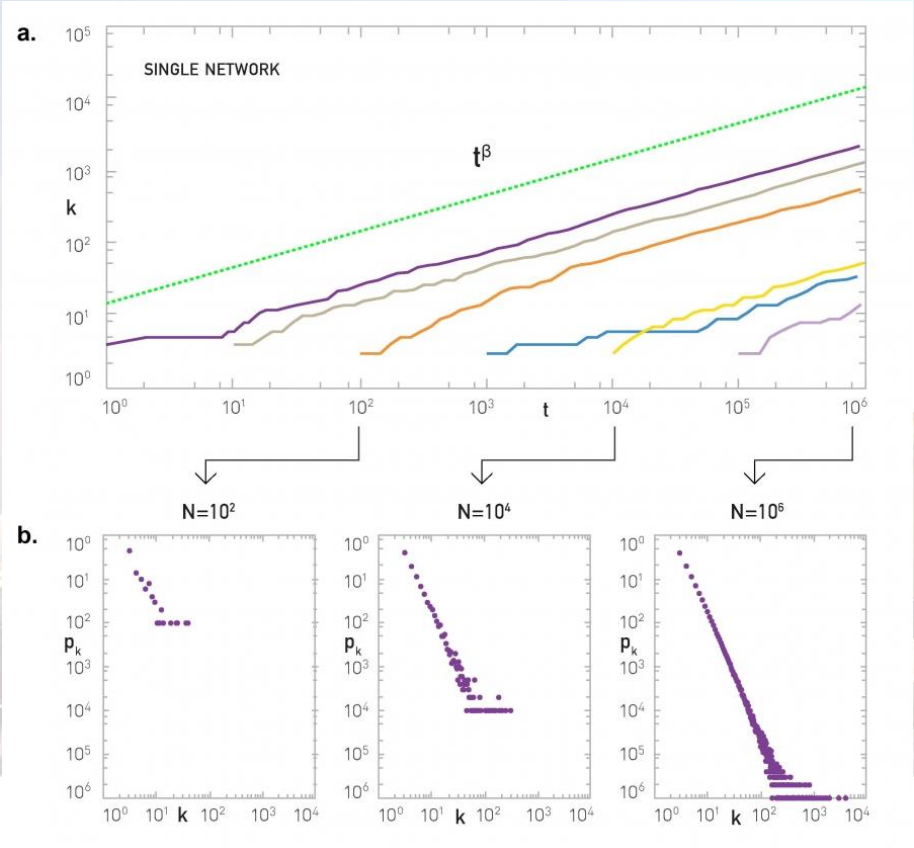
Degree Distribution over Time



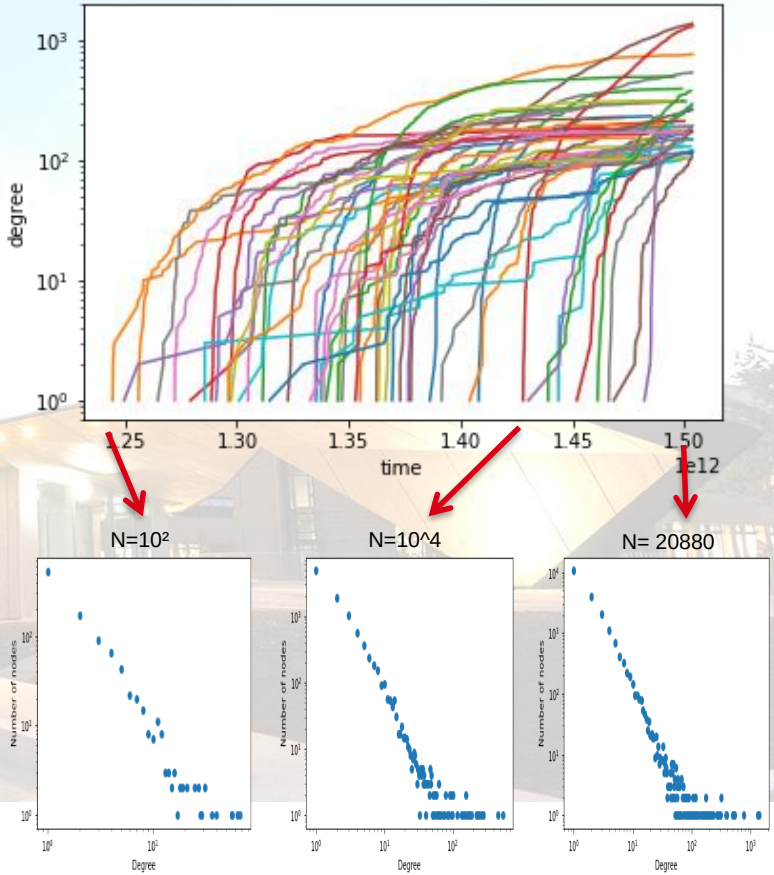
Filter k > 20



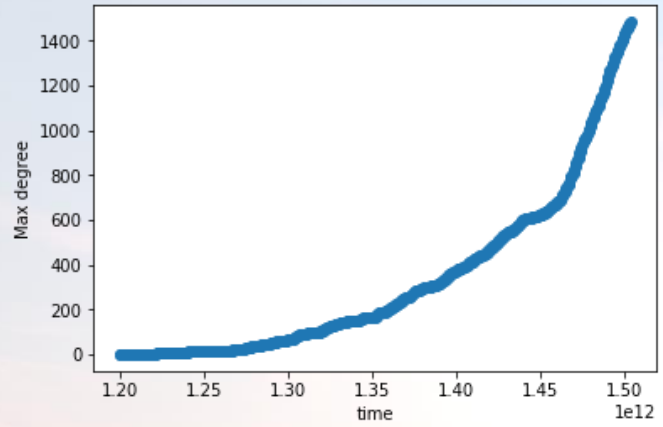
Degree Distribution over Time

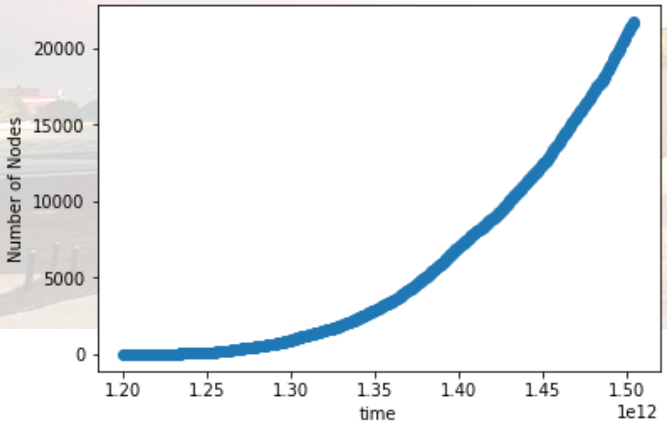


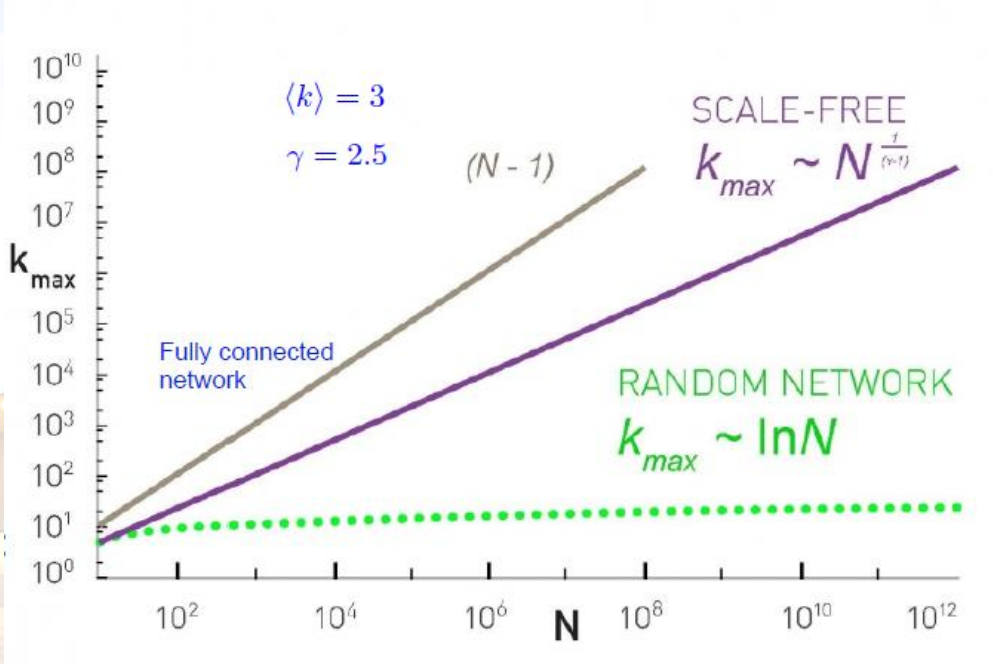
Filter k > 100



Evolution of k_{max}

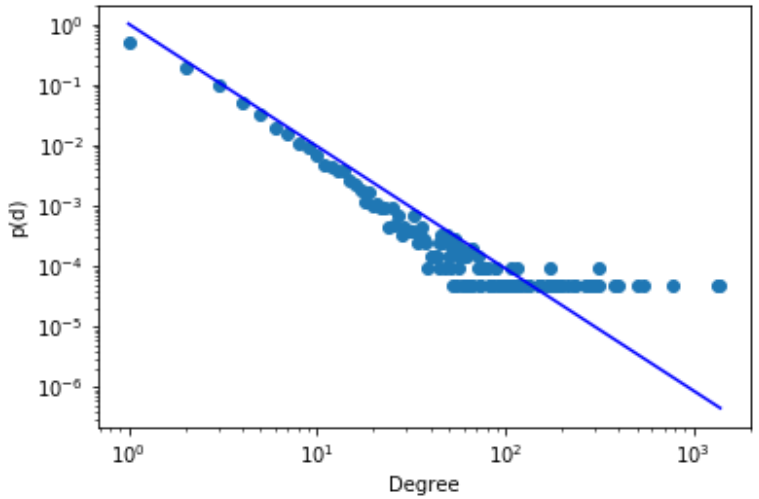


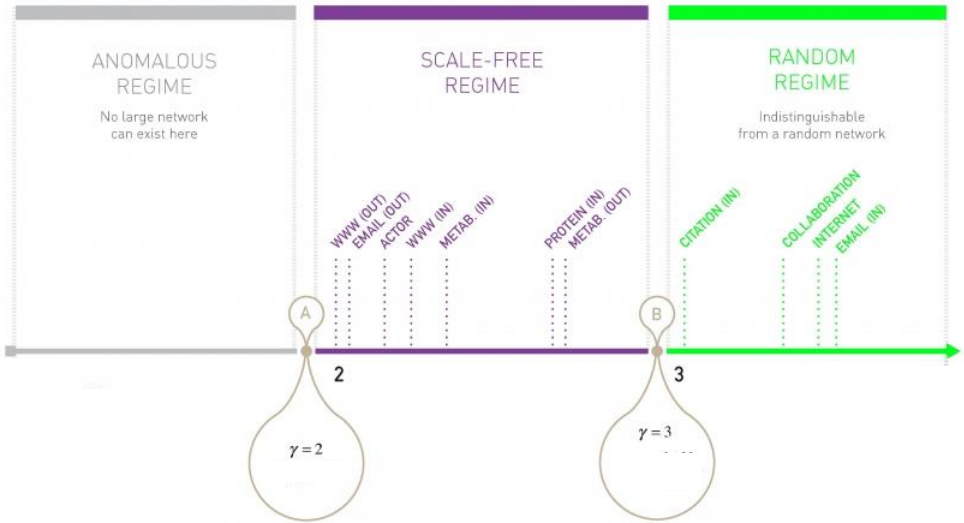




a ≈ 12.5

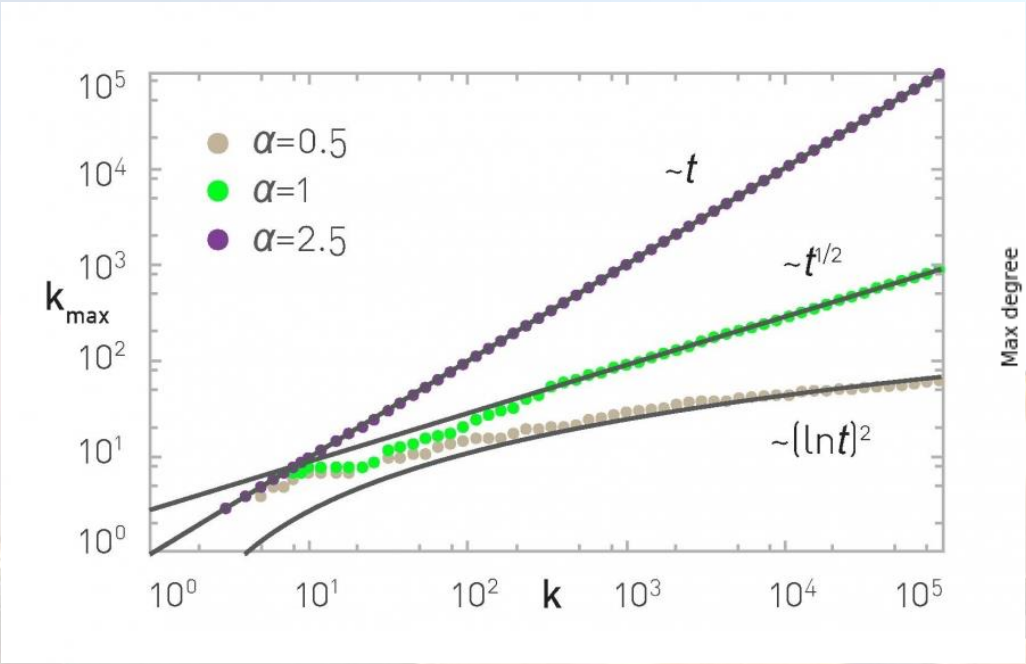
Preparing the network time analysis

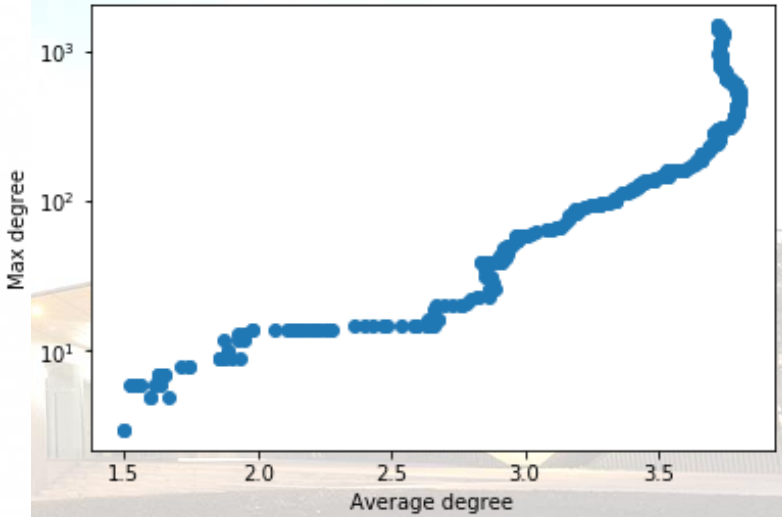




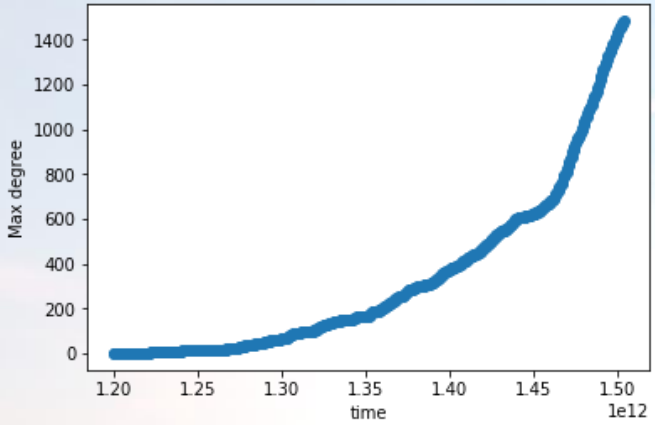
$\log p_{k\_in} - \gamma \log k_{in}$ $\gamma = 2.02$

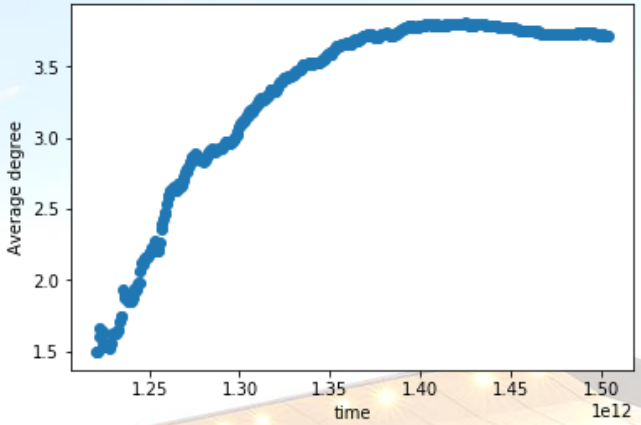
Preparing the network time analysis

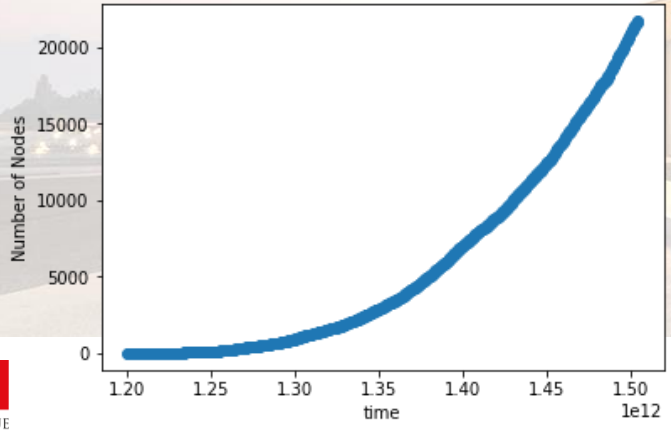


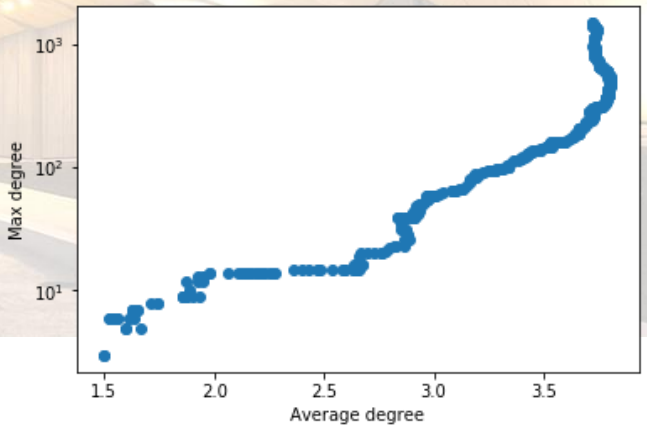


Network Evolution



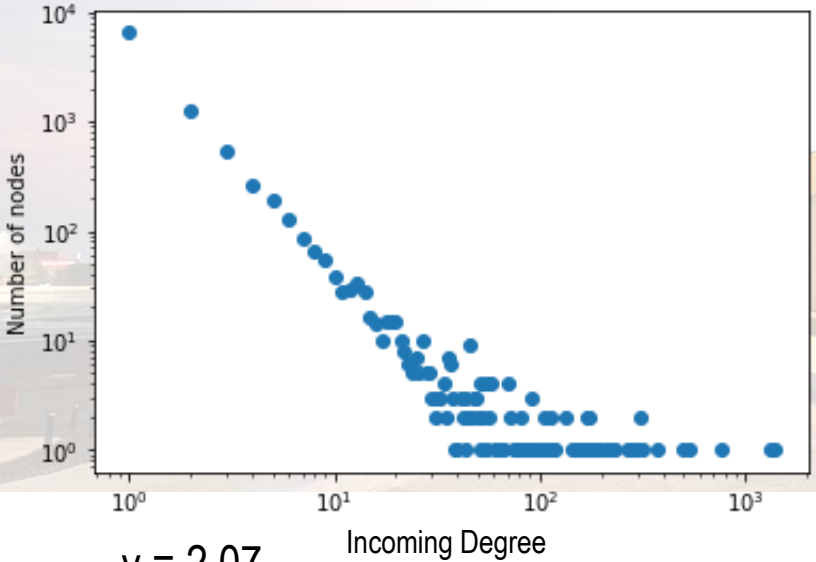






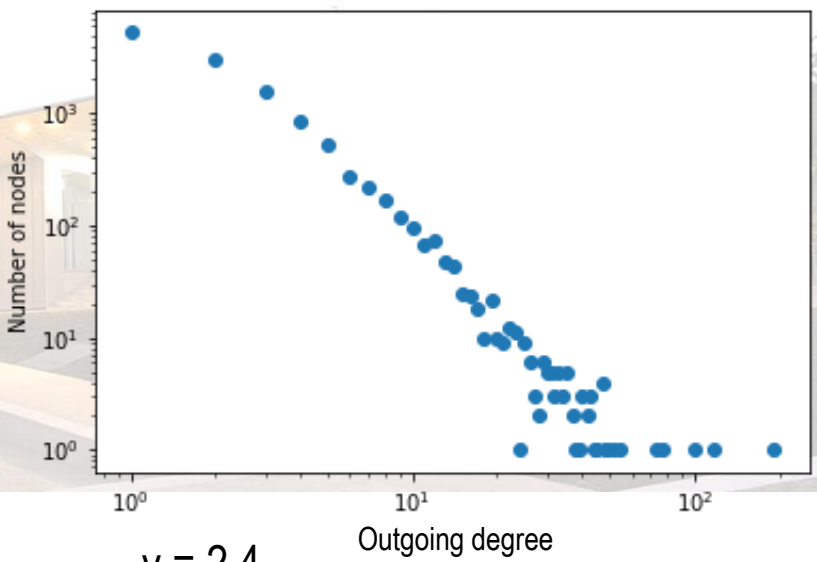
Directed Graph Interpretation – Final Degree Distribution

Incoming : Posting an Answer to a Question



γ = 2.07

Outgoing : Receiving an Answer for a Question



γ = 2.4

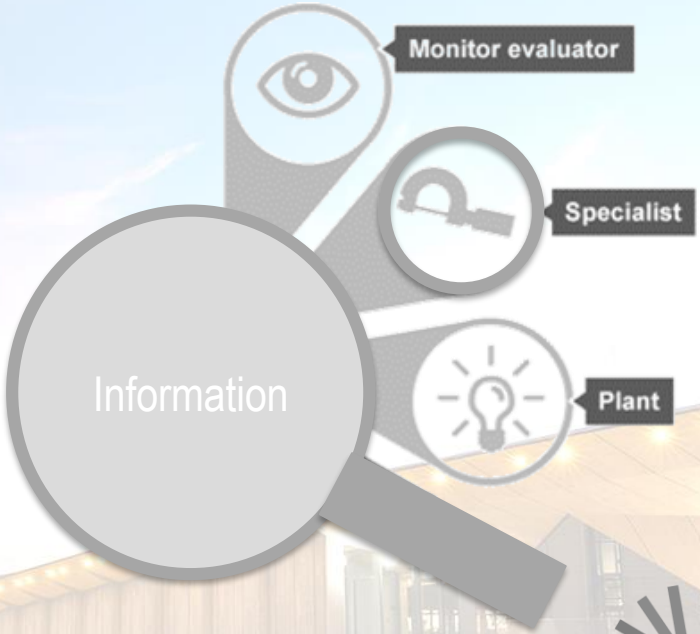
Team Roles


Resource investigator
Monitor evaluator
Teamworker
Specialist
Belbin team roles
Coordinator
Plant
Shaper
Implementer
Completer finisher
Thinking
Action
People

Team Roles

stackoverflow

Problem

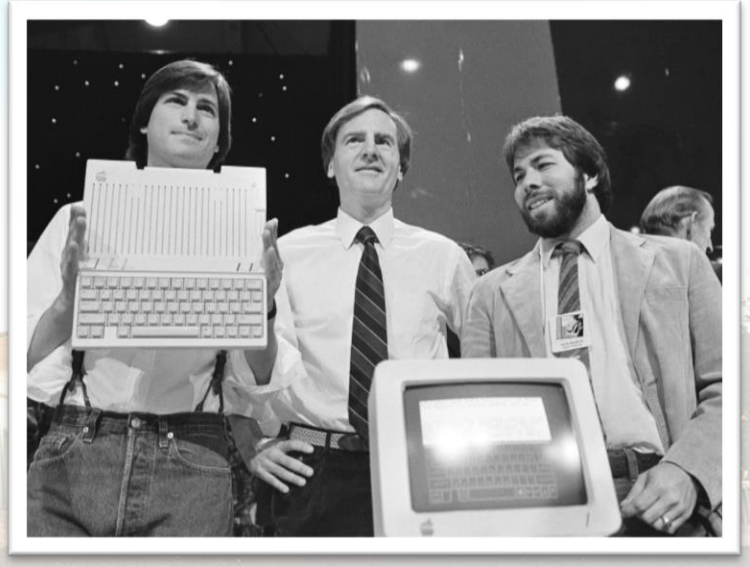


stackoverflow

Experts are hubs of the community


Area
Oakland
San Francisco
Alameda
San Leandro
Daly City
Hayward
Pacifica
San Mateo
Bair Island
El Granada
Half
Moon Bay
Palo Alt
Lobitos
La Honda


Area
San Francisco
Oakland
Cal
Tassajara
Daly City
Hayward
Pleasanton
Pacifica
Union City
Sunol
San Mateo
Fremont
Bair Island
El Granada
Half
Moon Bay
Don Edwards
National
Wildlife...
hp
NASA
Milpitas
STANFORD
UNIVERSITY
Lobitos
intel
La Honda
Campbell




Area
San Francisco
Oakland
Cal
Tassajara
Daly City
Hayward
Pleasanton
Pacifica
Union City
Sunol
San Mateo
Fremont
Bair Island
El Granada
Half Moon Bay
Milpitas
STANFORD UNIVERSITY
Lobitos
La Honda
Campbell
hp
NASA
YAHOO!
intel
ebay


WHO'S BEHIND THE CARTOON MAYHEM? BEING OBAMA
TIME
CAN WE TRUST
Google
WITH OUR SECRETS?
An exclusive inside look at the $100 billion empire that is dominating the Internet
BY ADI IGNATIUS


Thank you!

Sources

- Lecture Slides / Network Science Albert-László Barabási
- Silicon Valley - Google Maps, Flickr
- Meredith Belbin: Management Teams: Why They Succeed or Fail (1981)
  - https://media.licdn.com/mpr/mpr/AAIA_wDGAAAAAQAAAAAAAAufAAAAJGI1ZTI3OGUzLTA5NjEtNDdiYi1iNjc5LWE0ZTkyZWViMjE3MQ.png
- http://www.zonua.ie/articles/larry_page_sergey_brin_fashion_icons.php
- http://uk.businessinsider.com/steve-jobs-meeting-techniques-2015-8?r=US&IR=T
- Batman (https://i.ytimg.com/vi/KZqQUdFTvj0/maxresdefault.jpg)
- https://www.automobil-produktion.de/

Note: The presented values and plots might not match with the ones in the notebook, because we found and fixed a bug in our modules and used the corrected versions for the presentation.