CITEMAP

Mayank Kedia,   Kris Kooi,   Chrisantha Perera,   Flip Tanedo
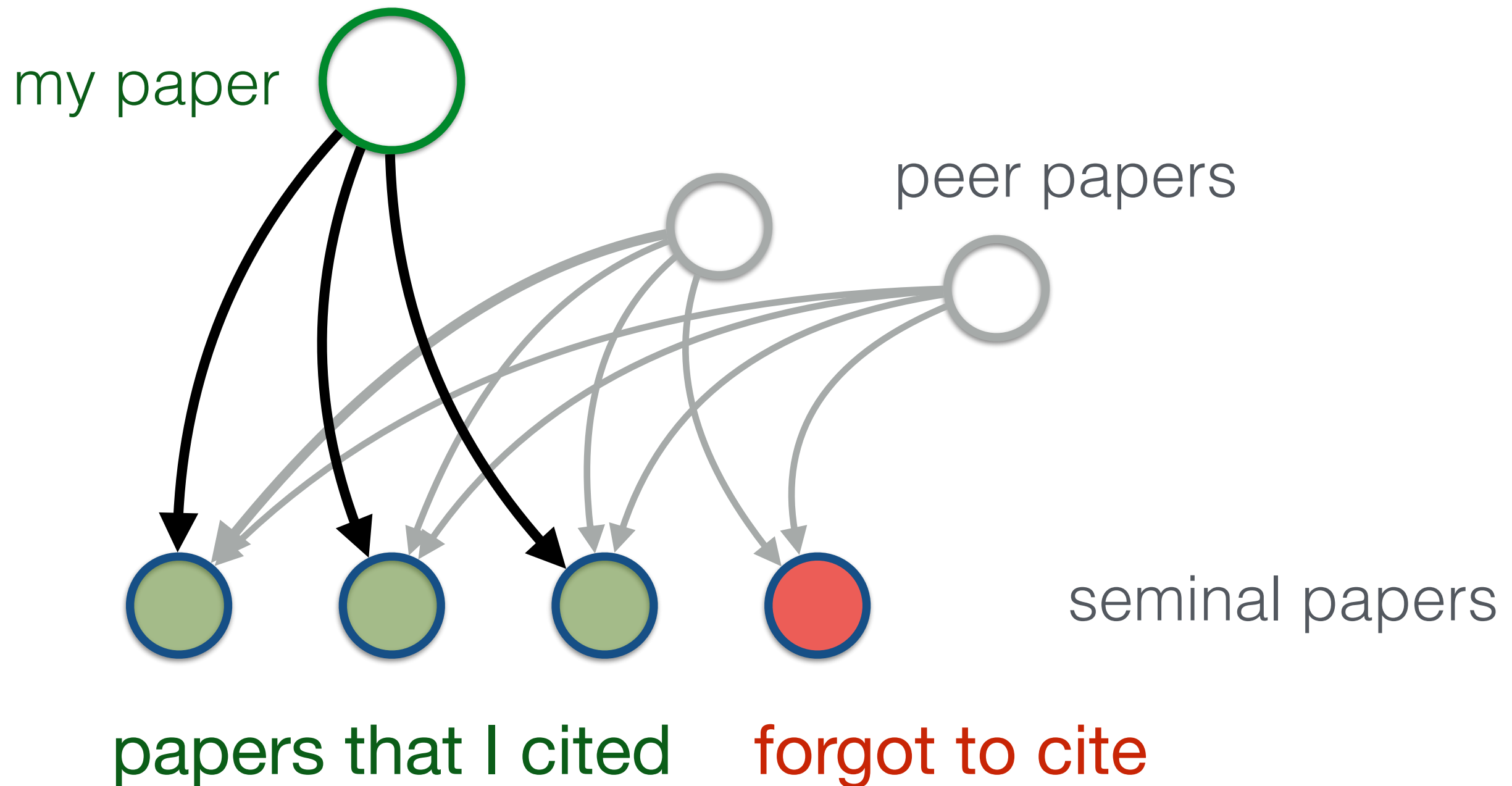
git.io/vWatB

SCIENCE HACK DAY
San Francisco 2015

The Problem

my paper

peer papers

papers that I cited   forgot to cite

seminal papers

The Data Set

Open access literature database for high energy physics

Clean: Remove empty data, incomplete data (old papers)
Make tractable: only take papers after 2000

~ 500,000 entries

```
{
    "abstract": {
        "summary": "We complete the effective potential calculation of the two-loop, top/bottom Yukawa
corrections to the Higgs boson masses in the Minimal Supersymmetric Standard Model, by computing the O(at^2
+ at*ab + ab^2) contributions for arbitrary values of the bottom Yukawa coupling. We also compute the
corrections to the minimization conditions of the effective potential at the same perturbative order. Our
results extend the existing O(at^2) calculation, and are relevant in regions of the parameter space
corresponding to tan(beta) >> 1. We extend to the Yukawa corrections a convenient renormalization scheme,
previously proposed for the O(ab*as) corrections, that avoids unphysically large threshold effects
associated with the bottom mass and absorbs the bulk of the corrections into the one-loop expression. For
large values of tan(beta), the new contributions can account for a variation of several GeV in the lightest
Higgs boson mass."
    },
    "recid": 618609,
    "title": {
        "title": "On the two loop Yukawa corrections to the MSSM Higgs boson masses at large tan beta"
    }
}
```

Possible Directions

In Vivo testing with academics

PubMed/CiteSeer^x dataset

FullText Analysis
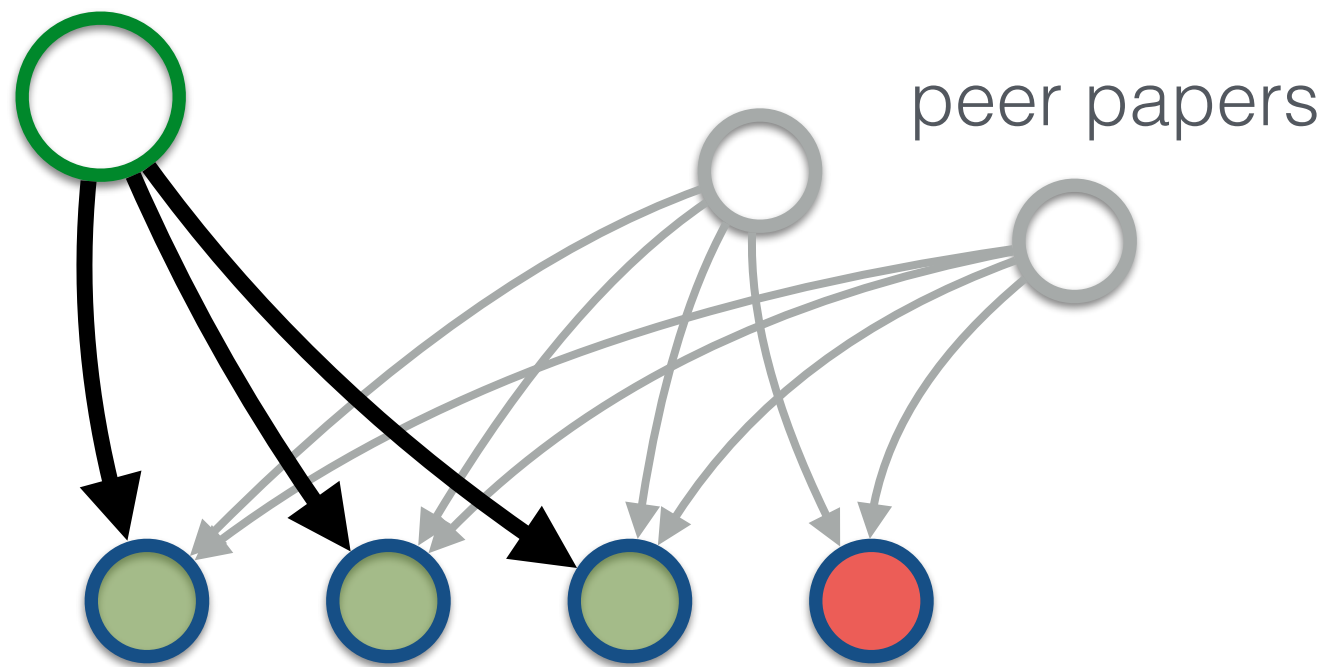PARTIAL DATASET: bit.ly/1RvuK4F

Heroku web service
Run on Spark

... and continue to spread & hack

SCIENCE HACK DAY

thanks everyone!

Technique

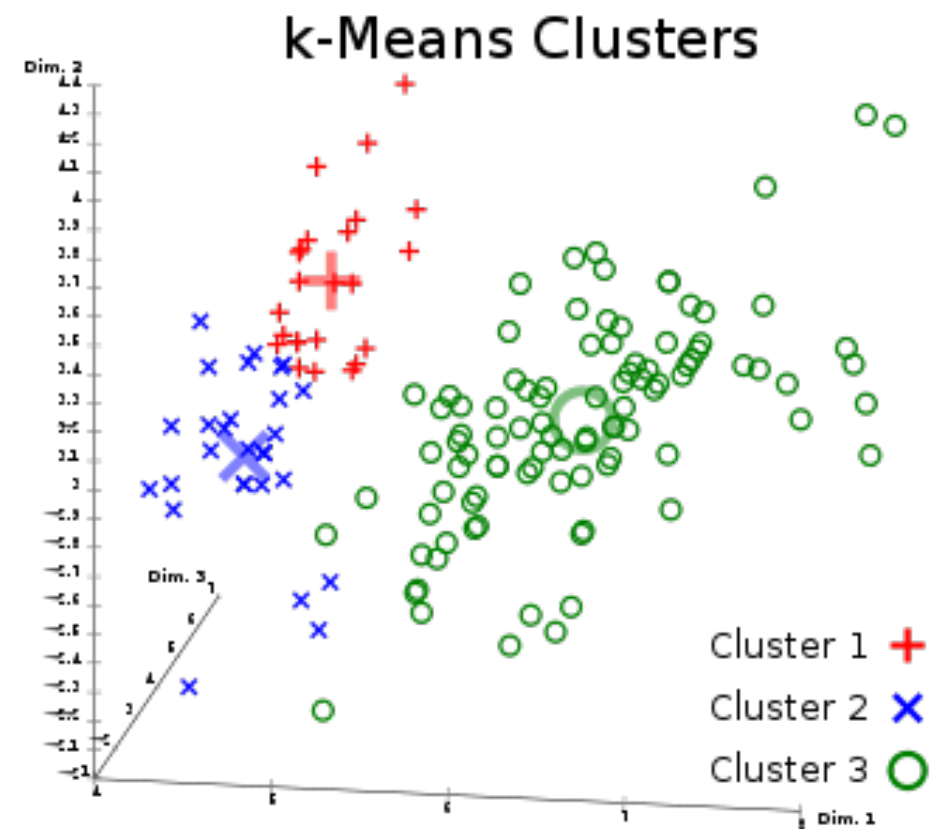## Using Peer Papers
RULES-BASED

peer papers

Bag of Words
K-MEANS CLUSTERING

k-Means Clusters

Cluster 1 +
Cluster 2 ✕
Cluster 3 ○

Wikimedia, Chire, *k-means*