



# Recommending Investors To Startups

**Brendan Wong**

*E-mail: [brendan.wong23@gmail.com](mailto:brendan.wong23@gmail.com)*

*Linkedin: [www.linkedin.com/in/brendan-ys-wong](https://www.linkedin.com/in/brendan-ys-wong)*

*Github: <https://github.com/brendan-ys-wong>*



## Who am I?



- Investment manager at Wong Centennial Foundation
- Chartered Accountant and Certified Professional Accountant since 2012
- Previously worked at Netflix as Senior Business Analyst
- Prior to Netflix, worked at Ernst & Young and KPMG
- Passionate about machine learning and artificial intelligence investing!



# Landscape

## Dataset

- 139K startup investment transactions from 1999-2015 provided by Crunchbase

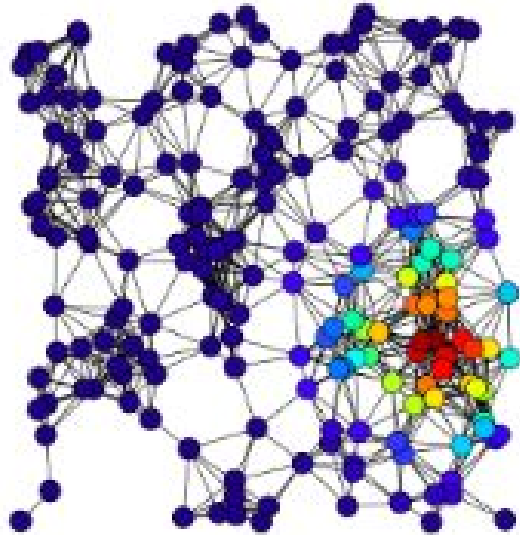
## Matchmaking problem - too many choices/inefficient resource allocation

- 20K startups, 15K investors
- 400+ investors with at least 50 investments

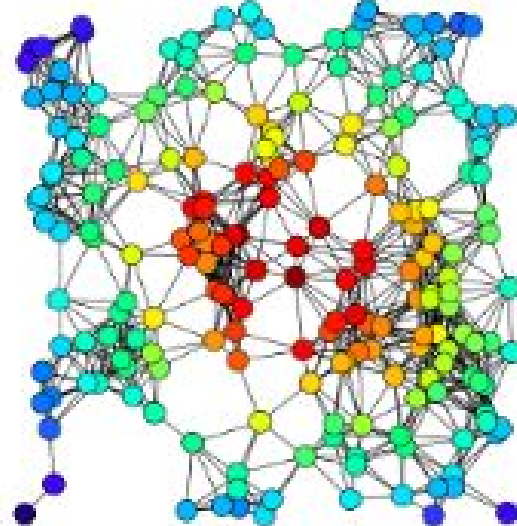
*Tough to find right investor....great machine learning opportunity!*

# Exploratory analysis on startup network

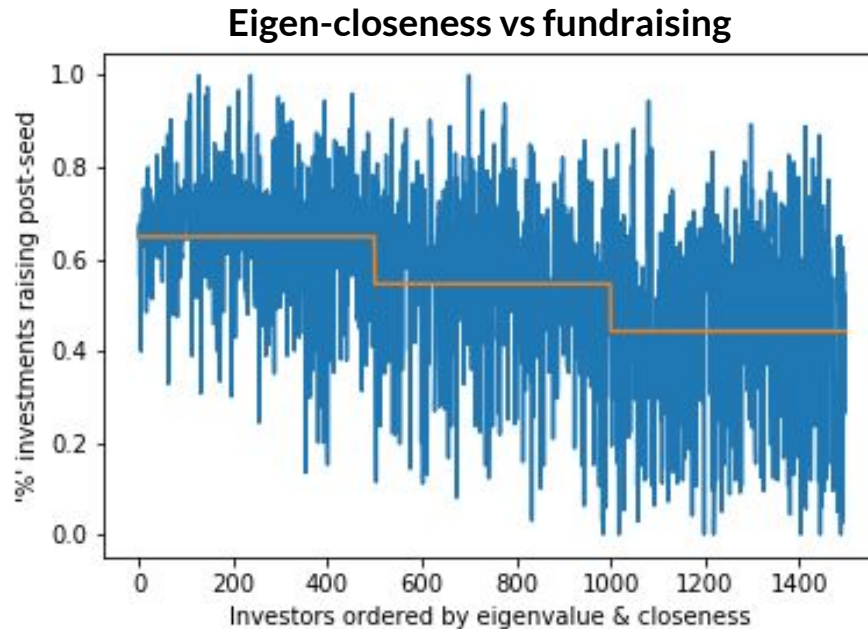
Eigenvector (*connections*)



Closeness (*Information*)



# Interpreting the signal





## First model: investor similarity model

Identifying similar investor based on investing history using jaccard similarity metric

*Example:*

- *Startup A: Seed Investor #1, Venture Investor #1*
- *Startup B: Seed Investor #1, Venture Investor #1, Venture Investor #2*

*Seed Investor #1 and Venture Investor #1 are similar since they co-invest frequently. Recommend Venture Investor #2 to Startup A.*



# First model: results

**Recall** - Ability to identify successful investor-startup pairs in held-out data set

**Precision** - Accuracy in those recommendations of investor-startup pairs

Of the two, recall is most strategically important.

The cost of a startup pitching to 30 investors instead of 10 is low. We care about if this model is finding signal in identifying successful investor-startup matches, recall is critical.

Baseline recall - 8.09% (*Recommending the largest investors*)

Model recall - 12.82% (*58% improvement on baseline*)



## Second model: expanding the scope

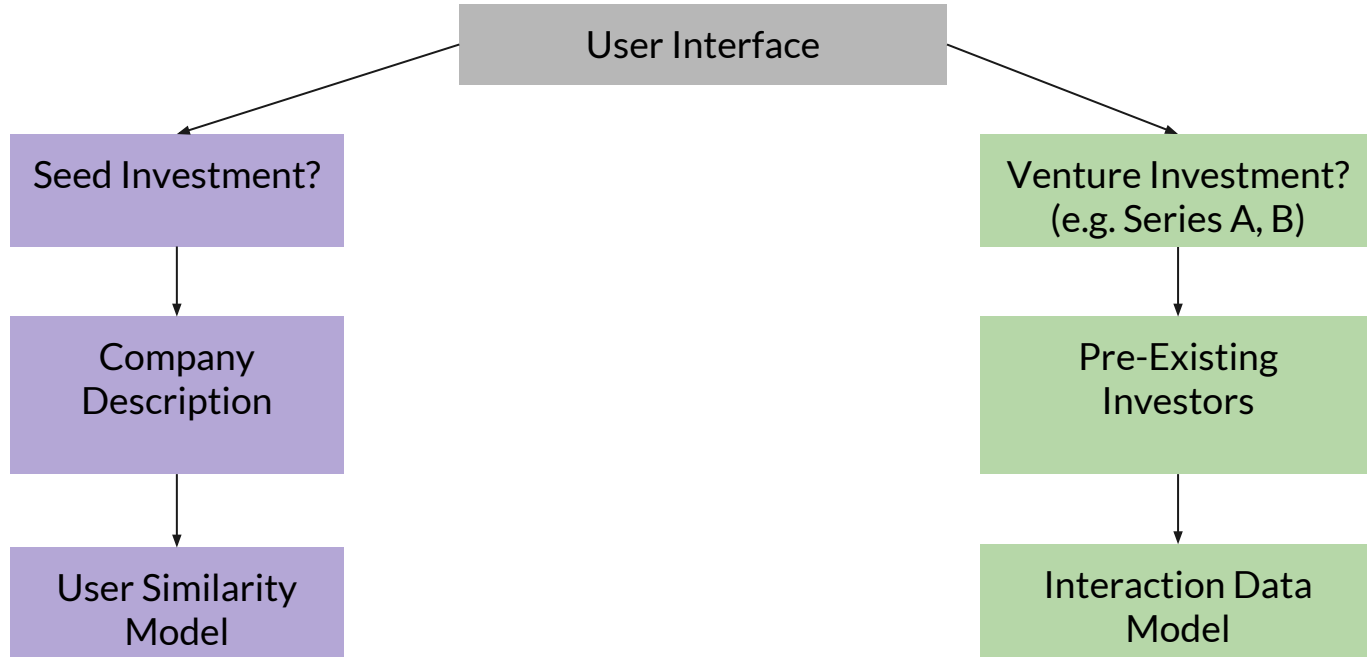
For new companies without pre-existing investors, identify similar companies based on analyzing company description using natural language processing

### Steps:

1. Tokenize company description
2. Vectorize tokens based on co-occurrence rates from a common-crawl corpus
3. Calculate cosine similarity between NewCo and all other companies
4. Return investor list for all sufficiently similar companies



# Recommendation System





## Take-aways & next steps

- Improving the fundraising process using machine learning is an iterative process
- Model provides a more diverse selection of investors likely to be interested in your start-up with a similar performance metrics compared with current approach
- Adding in personnel information (e.g. University, past career experience) would be next step and likely further boost signal and quality of recommendation



# Thank you!

**Brendan Wong**

*E-mail: [brendan.wong23@gmail.com](mailto:brendan.wong23@gmail.com)*

*Linkedin: [www.linkedin.com/in/brendan-ys-wong](https://www.linkedin.com/in/brendan-ys-wong)*

*Github: <https://github.com/brendan-ys-wong>*