

Getting an Edge with Network Analysis in Python

A Gentle Introduction

Alon Nir
(@alonnir)

PyCon US
May 2021

About Me

- Senior data scientist at Spotify

Note: this talk is separate to my work at Spotify, and I do not represent my employer at PyCon. Previous iterations of this talk were given at [PyCon UK](#) and [PyData Berlin](#).

- London based
- Sometimes I tweet interesting things: @alonnir

Our Plan For Today

What we'll do:

1. Introduce networks, network analysis and a few motivating examples.
2. Brief theoretical intro (the basic building blocks)
3. Go deeper with hands on analysis of a real network
4. Wrap up, next steps and some homework assignments



Our Plan For Today

What we won't do:

- In the interest of time we'll touch on just a few of the many (many!) things there's to learn, just enough to get you started.
- No mathematical formulations! No matrix representations! No Greek letters!
- Dive deep into theory (results are often intuitive but the algorithms to get them are not.)



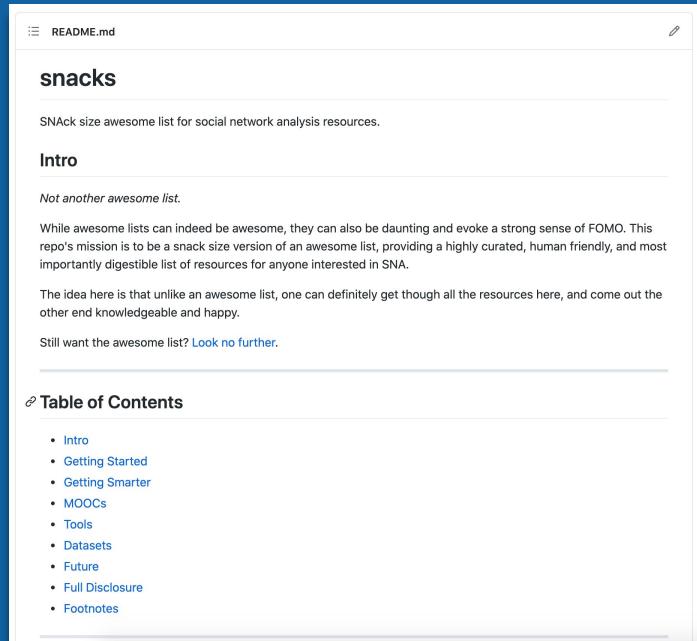
Would You Like to Know More?

Due to time limitations we'll barely scratch the surface.

SNAcks (snack size awesome list for social network analysis resources) is a repo with carefully curated resources for people who want to get started with social network analysis.

Check it out!

<https://github.com/alonnir/snacks>



The screenshot shows the content of a README.md file for a GitHub repository named 'snacks'. The file is titled 'snacks' and describes it as a 'SNAck size awesome list for social network analysis resources'. It includes an 'Intro' section with a note that it's a 'Not another awesome list'. The text explains that the repo's mission is to provide a highly curated, human-friendly, and digestible list of resources for anyone interested in SNA. It emphasizes that unlike a full awesome list, one can get through all the resources here and come out the other end knowledgeable and happy. A link 'Look no further.' is provided. Below the text is a 'Table of Contents' section with a list of links to various sections: Intro, Getting Started, Getting Smarter, MOOCs, Tools, Datasets, Future, Full Disclosure, and Footnotes.

```
snacks

SNAck size awesome list for social network analysis resources.

Intro

Not another awesome list.

While awesome lists can indeed be awesome, they can also be daunting and evoke a strong sense of FOMO. This repo's mission is to be a snack size version of an awesome list, providing a highly curated, human friendly, and most importantly digestible list of resources for anyone interested in SNA.

The idea here is that unlike an awesome list, one can definitely get through all the resources here, and come out the other end knowledgeable and happy.

Still want the awesome list? Look no further.

Table of Contents

- [Intro]
- [Getting Started]
- [Getting Smarter]
- [MOOCs]
- [Tools]
- [Datasets]
- [Future]
- [Full Disclosure]
- [Footnotes]
```

What Is a Network Anyway?

And why should we care?

points



points and lines



points and lines



If you can draw a conceptual relationship (line) between any two things (points), it's likely you can model them as a network.

Points	Lines
People	Friendships, mating, places of work, following, texting, money transfers...
Companies	Trade, employees move
Bus stops	Buses, passengers
Bicycle/scooter sharing schemes	riders
Countries	Trade, flight routes, Eurovision points
Wikipedia articles	Internal links
NBA teams	Player trades
Characters in a film/show/book	Scenes together, conversation
Academics	Paper co-authorships, citations

Obvious Example - a Social Network





POINT



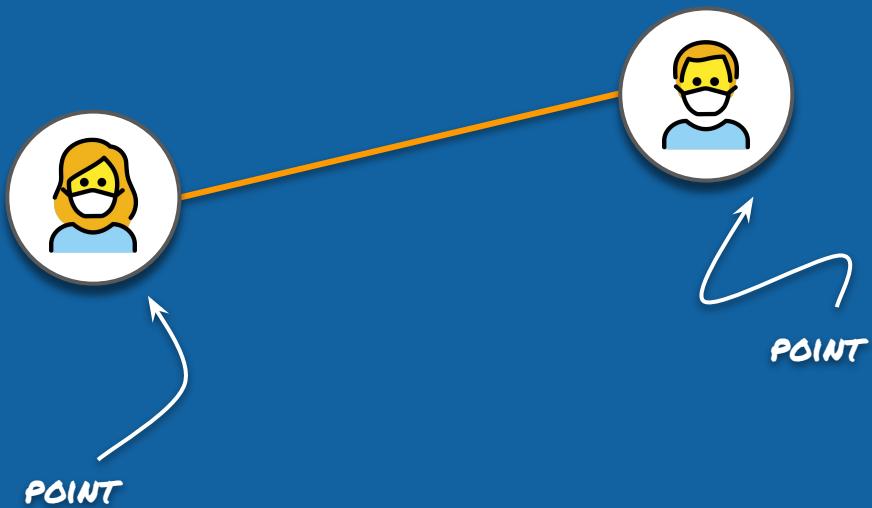
POINT

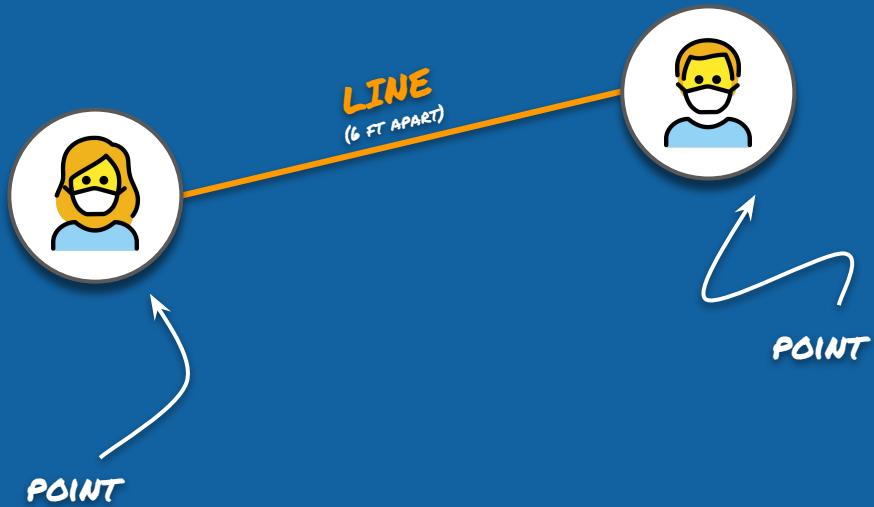


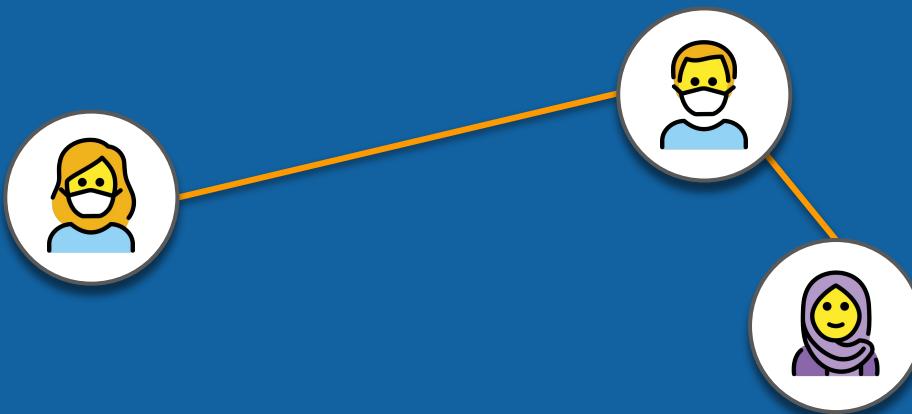
POINT

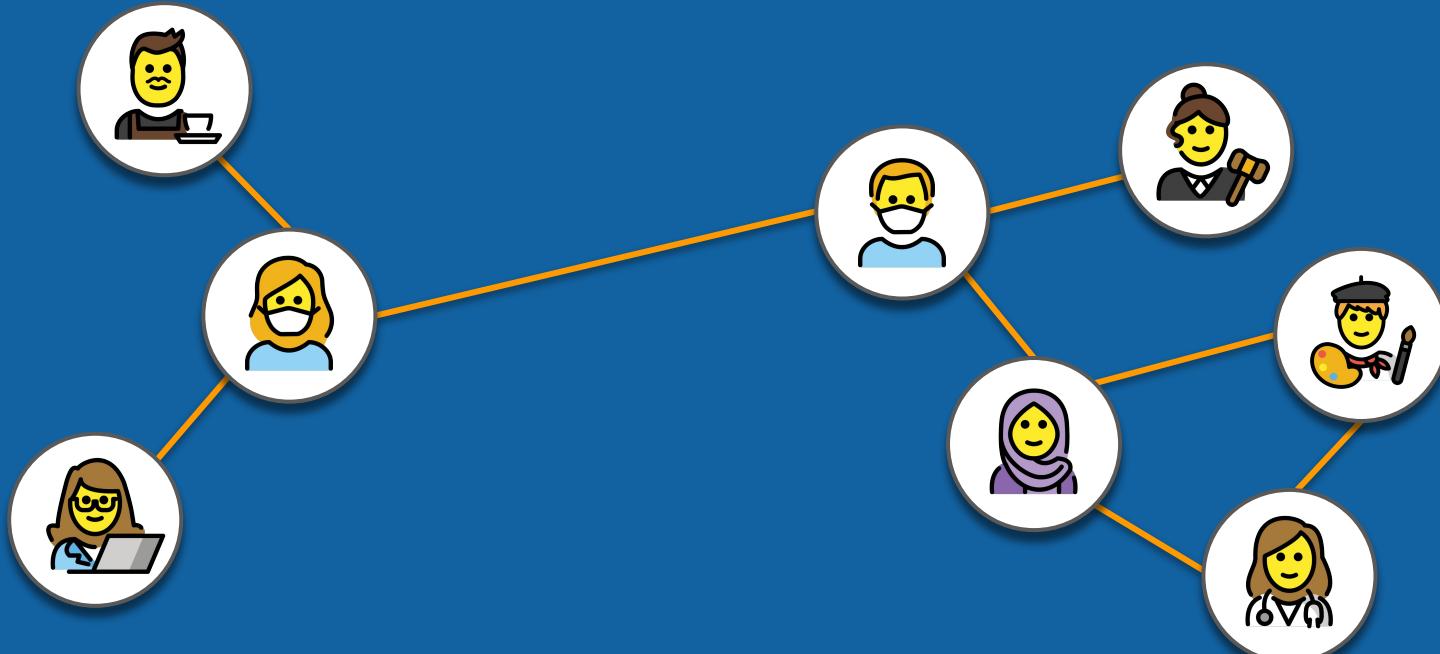


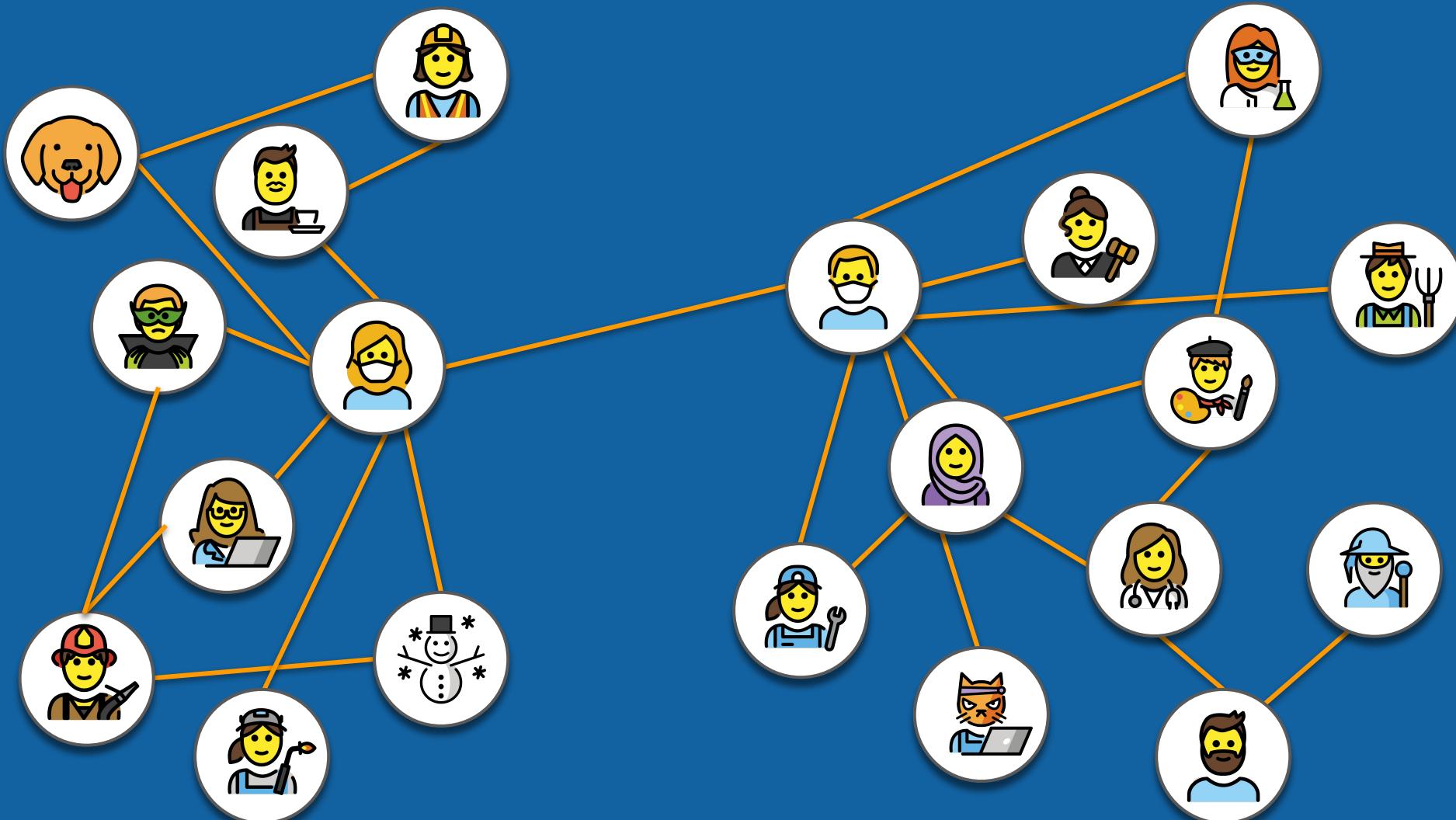
POINT











Points and Lines have more formal names, which vary by discipline

Points	Lines	Discipline
Nodes	Edges, Links	Computer Science
Vertices	Edges, Arcs	Math
Sites	Bonds	Physics
Actors	Ties, Relations	Sociology

Source:

Jean Mark Gawron, Python for Social Science,

https://gawron.sdsu.edu/python_for_ss/course_core/book_draft/Social_Networks/Social_Networks.html#what-are-networks

Points and Lines have more formal names, which vary by discipline



Points	Lines	Discipline
Nodes	Edges, Links	Computer Science
Vertices	Edges, Arcs	Math
Sites	Bonds	Physics
Actors	Ties, Relations	Sociology

Source:

Jean Mark Gawron, Python for Social Science,

https://gawron.sdsu.edu/python_for_ss/course_core/book_draft/Social_Networks/Social_Networks.html#what-are-networks

Main Types of Networks

UnDirected

Facebook
Linkedin

Connections are **directionless** and **reciprocal** - if A is friends with B on facebook, B is also friends with A.



Also: MultiGraph and Multi-Directed-Graph

Main Types of Networks

UnDirected

Facebook
Linkedin

Connections are **directionless** and **reciprocal** - if A is friends with B on facebook, B is also friends with A.



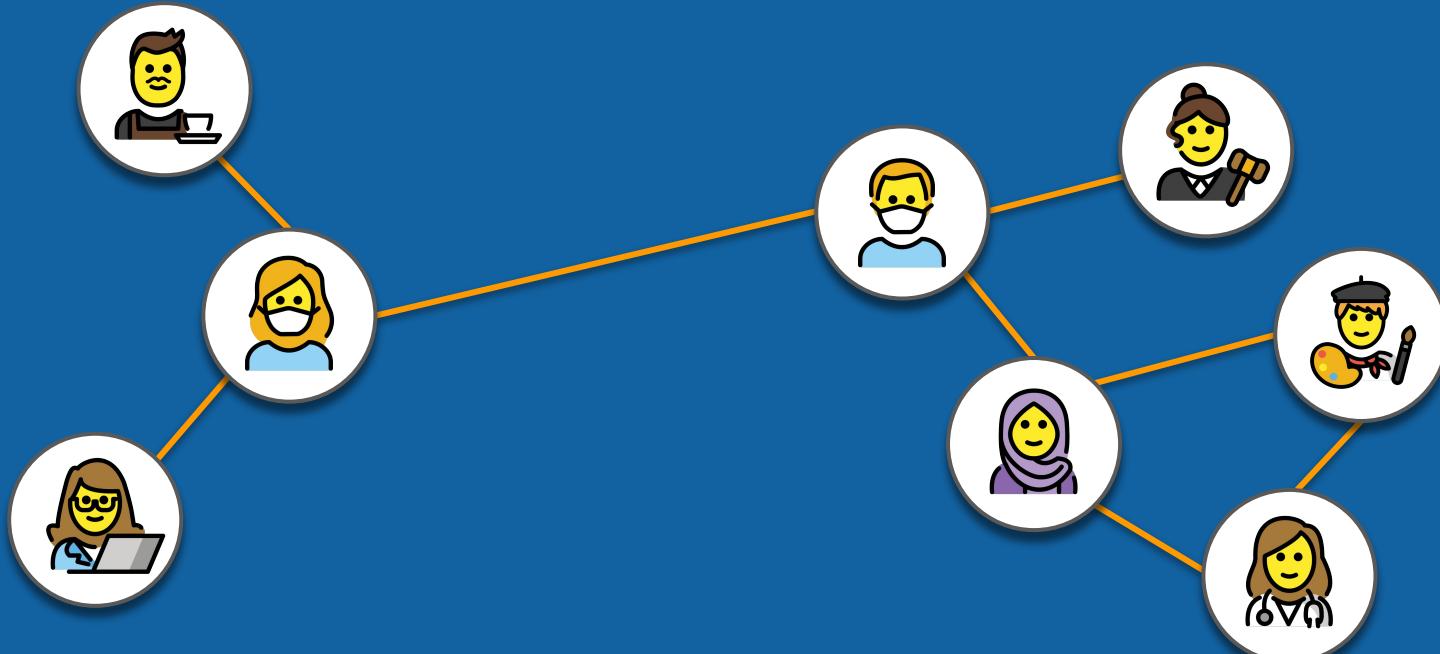
Directed

Twitter
Snapchat
Instagram
TikTok

Connections don't have to be reciprocal - if A follows B on Twitter, B does not necessarily follow A



Also: MultiGraph and Multi-Directed-Graph



NODE



NODE



NODE



NODE



NODE

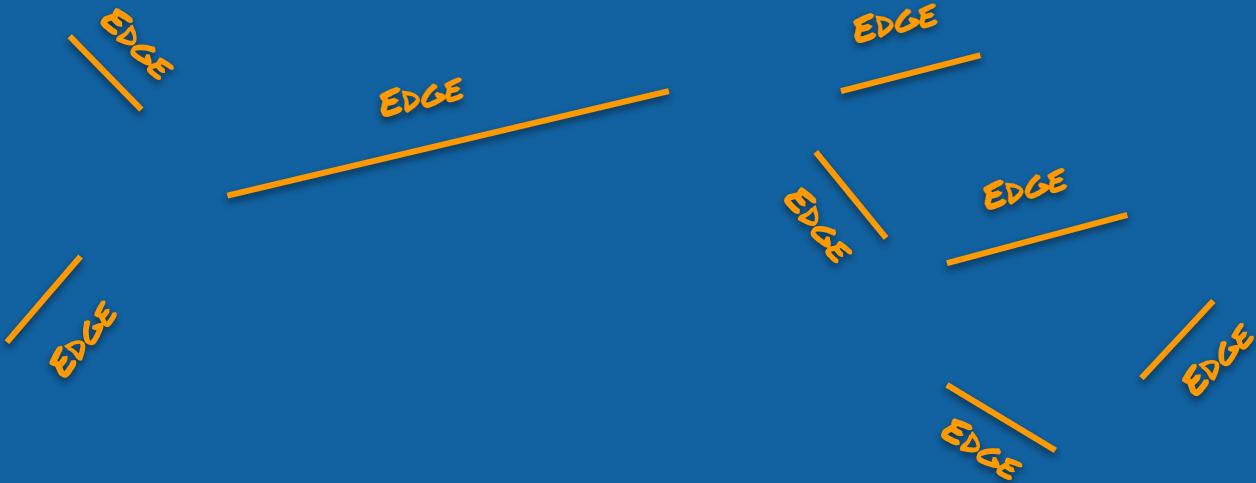


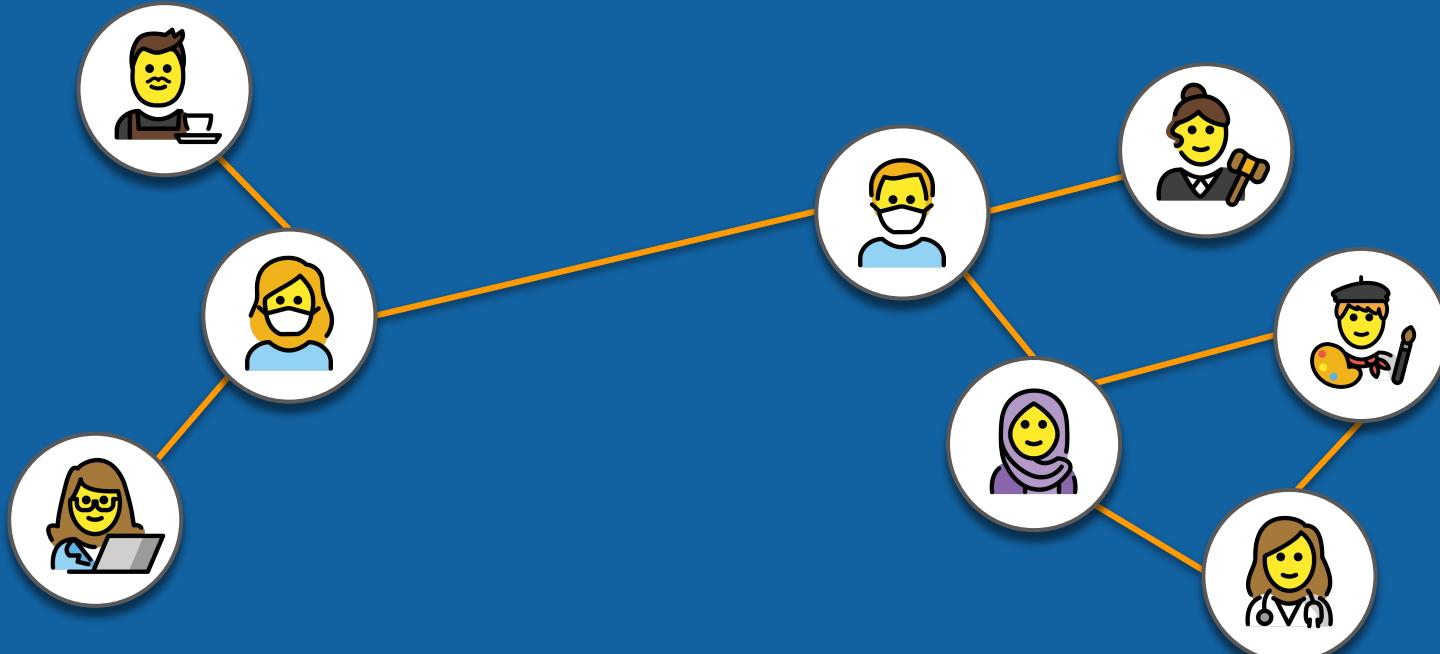
NODE

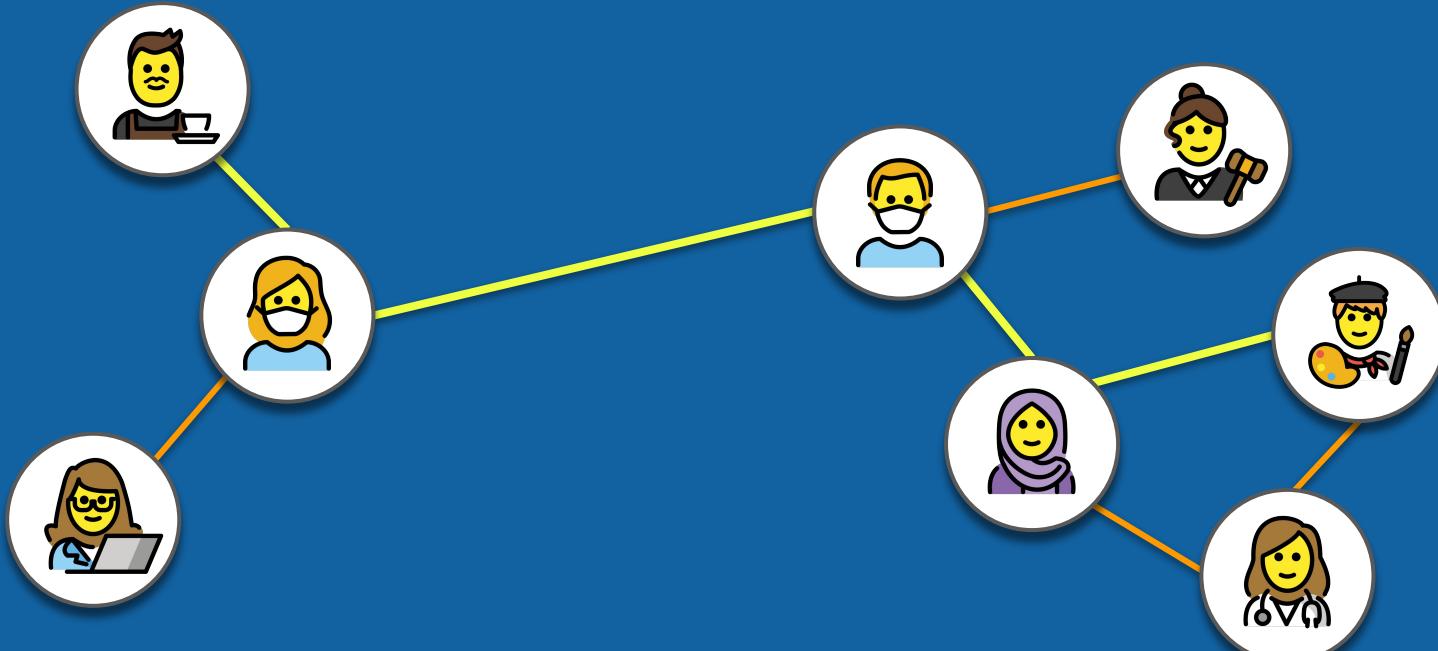


NODE



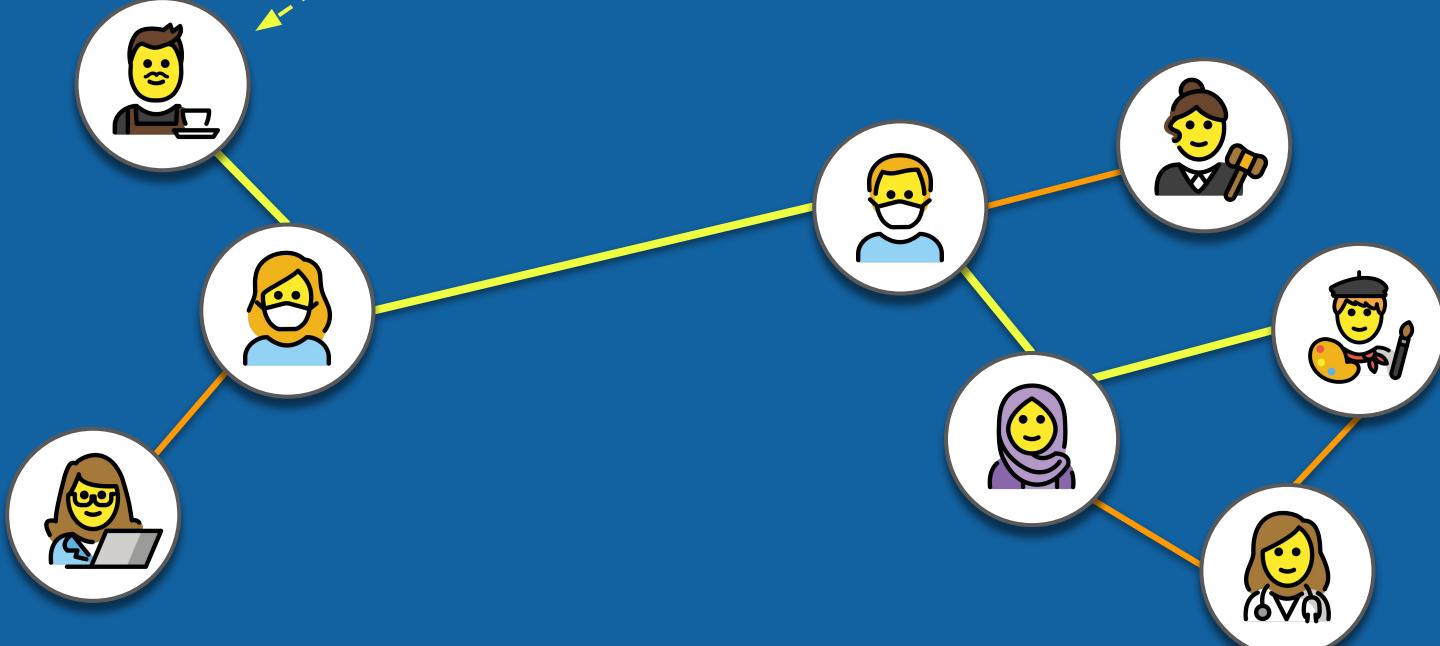






Path = edges from a node to another node, w/o repetition

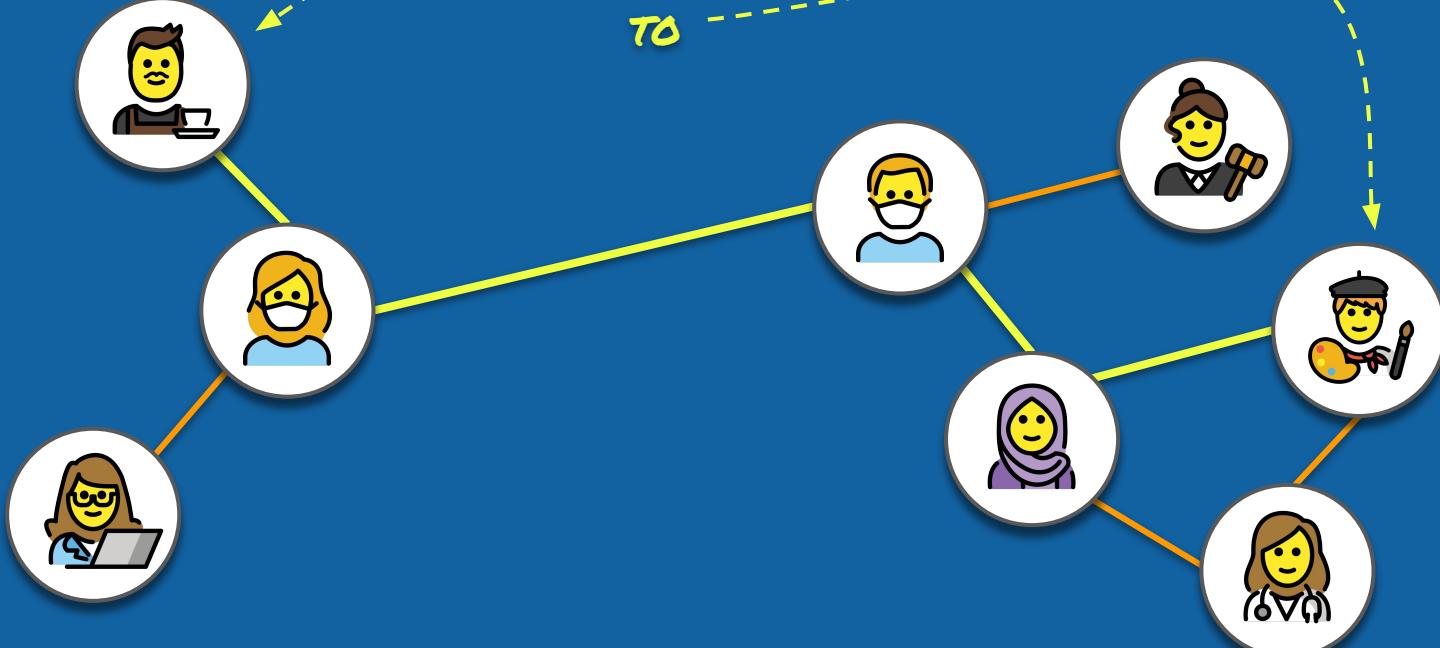
PATH
FROM



Path = edges from a node to another node, w/o repetition

PATH

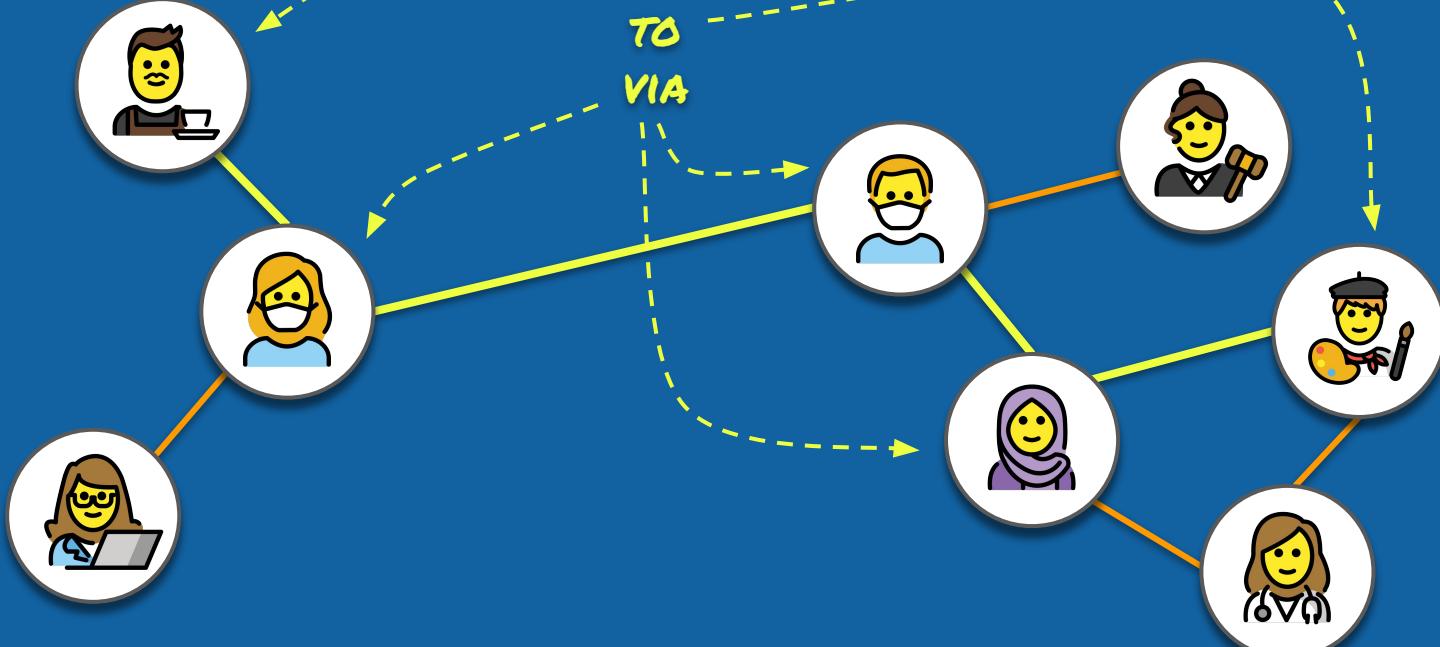
FROM
TO



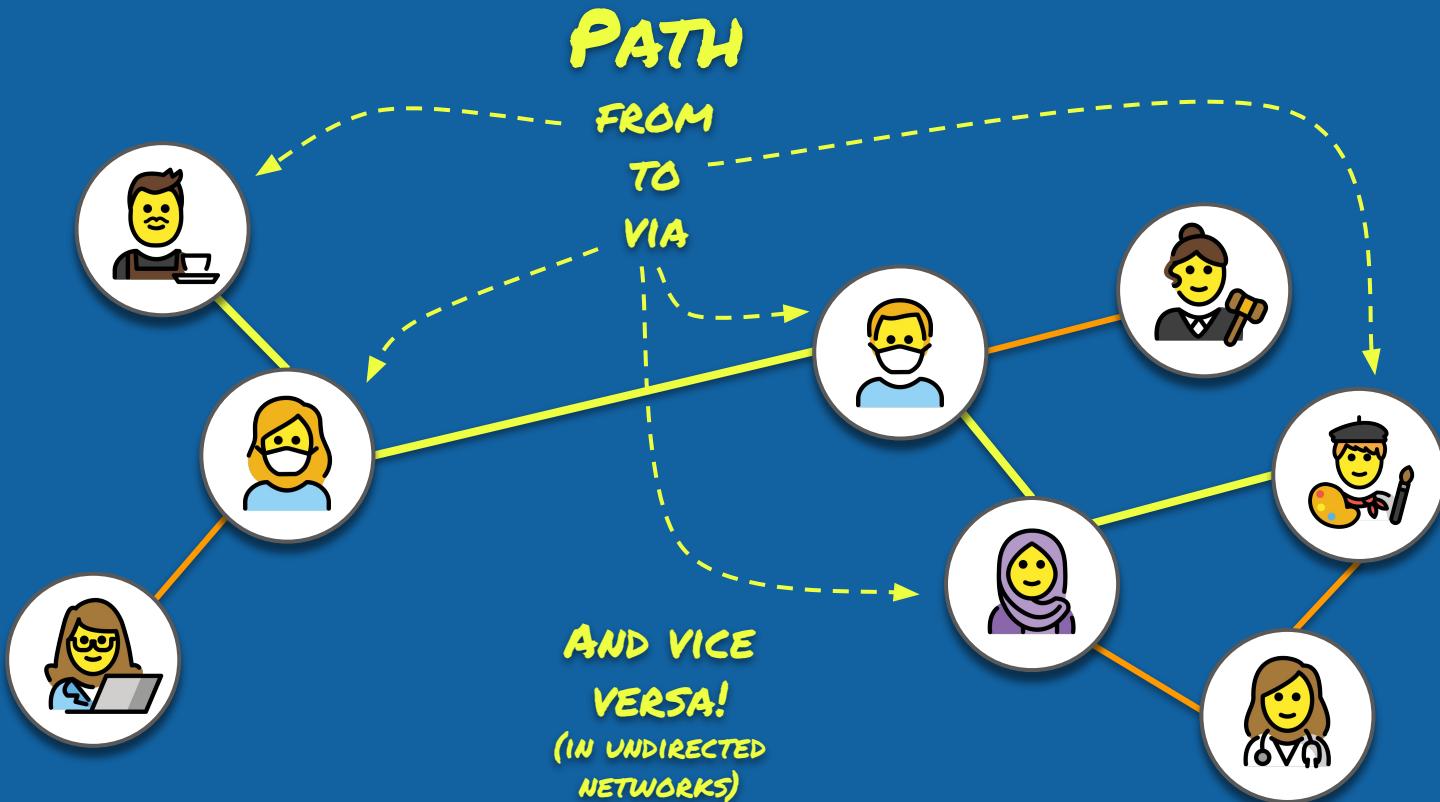
Path = edges from a node to another node, w/o repetition

PATH

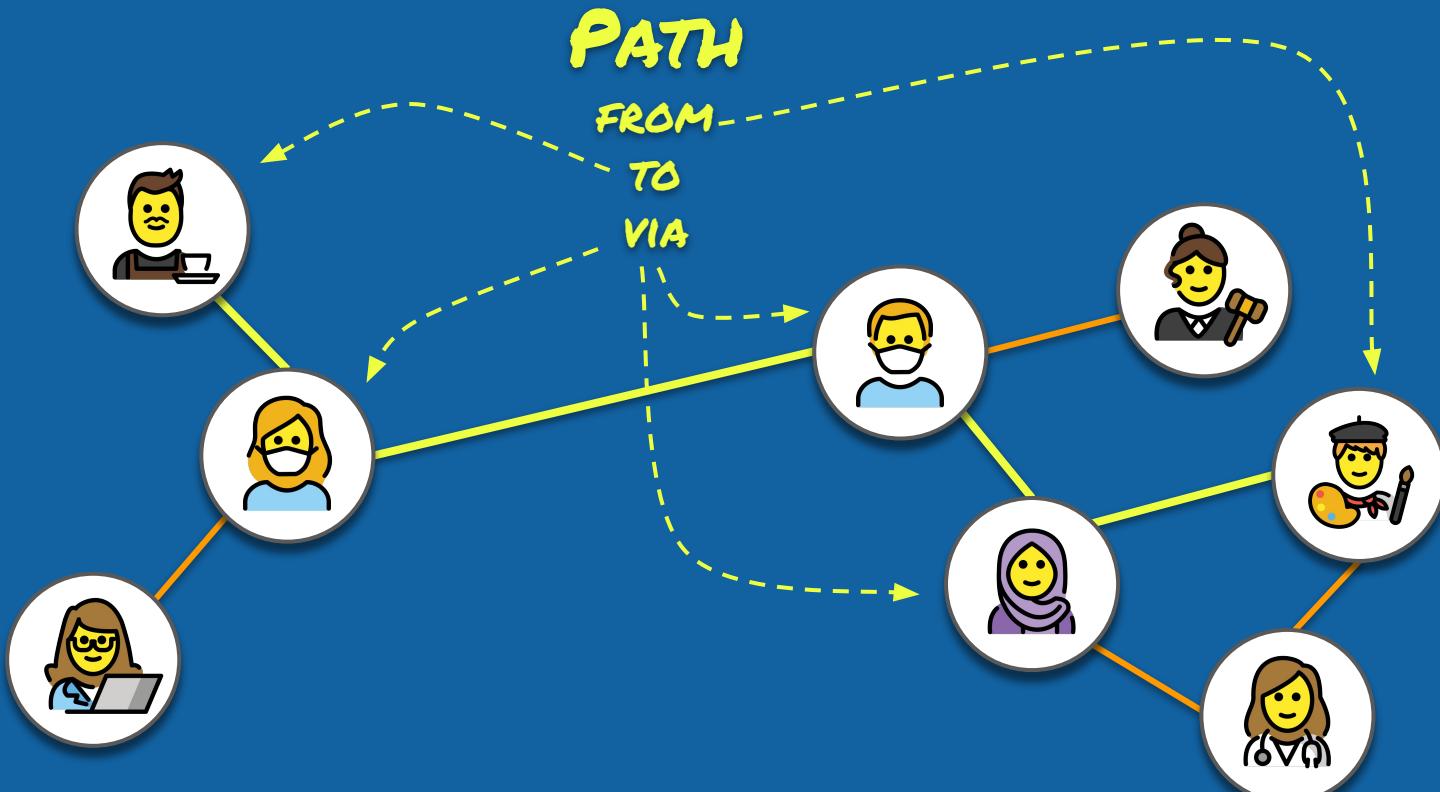
FROM
TO
VIA



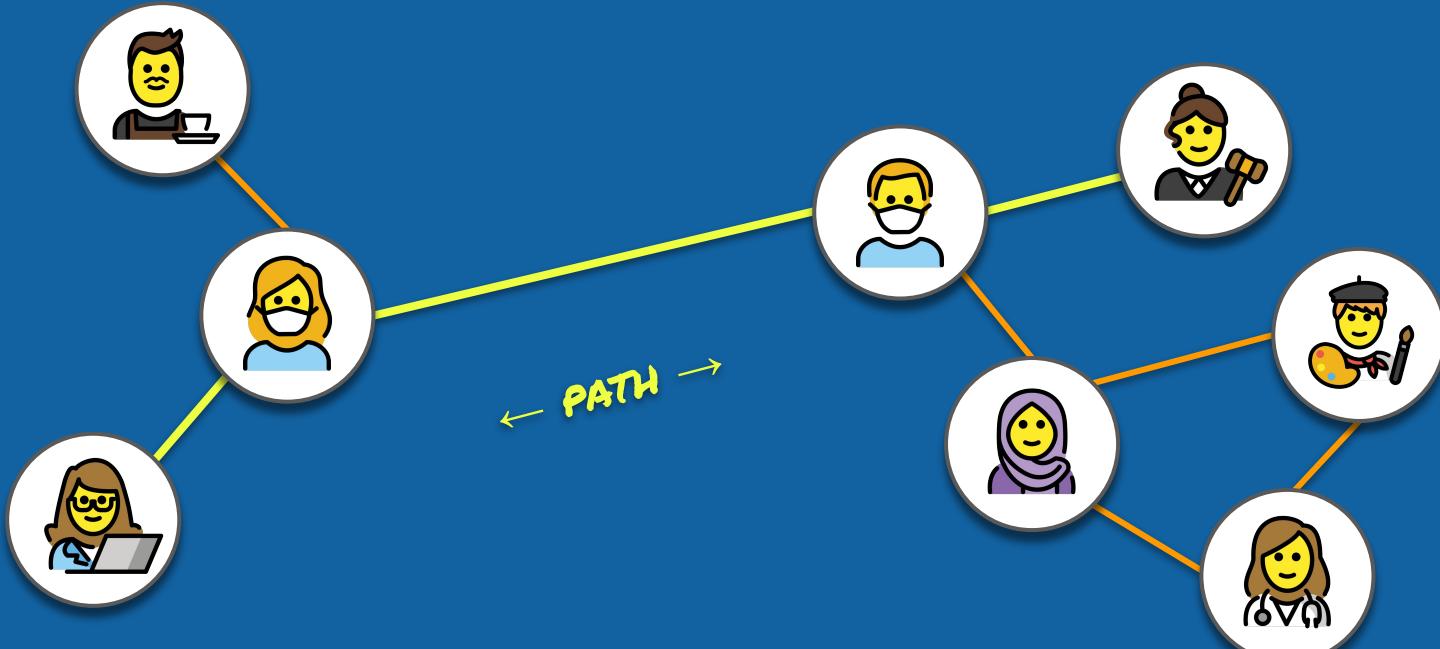
Path = edges from a node to another node, w/o repetition



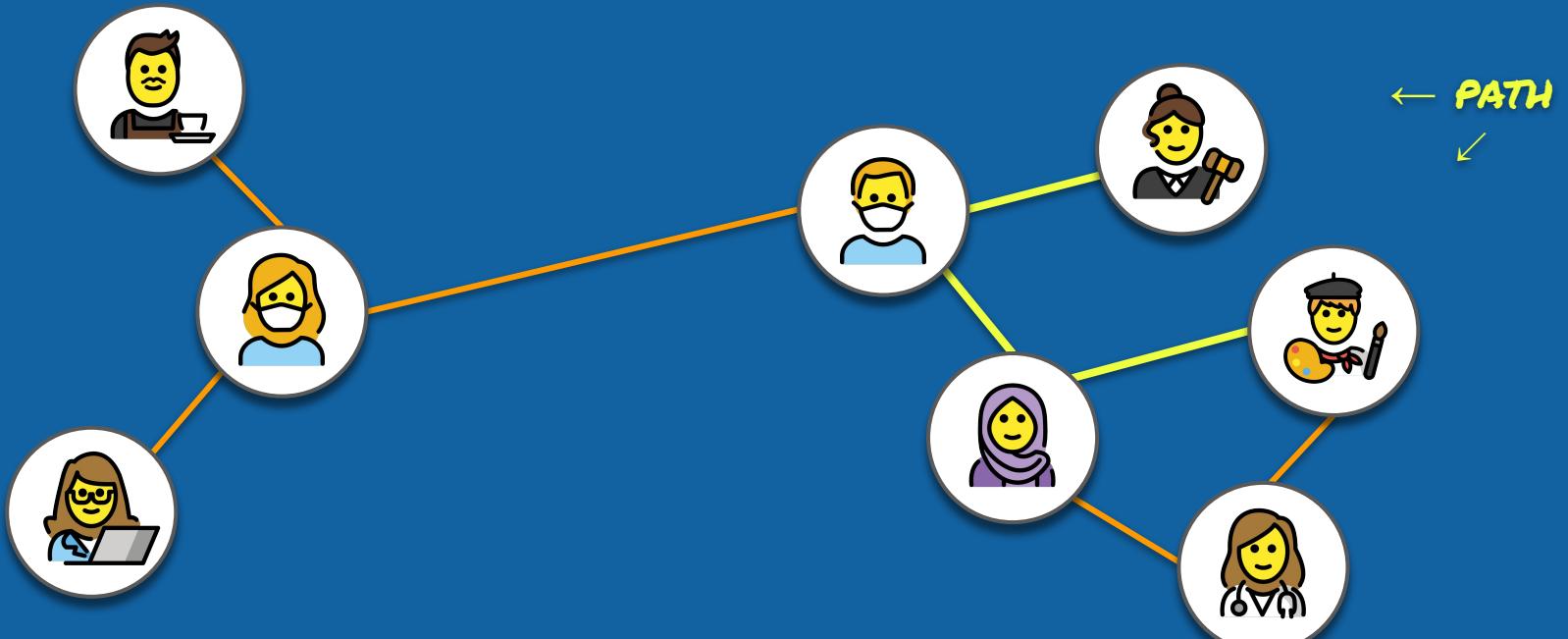
Path = edges from a node to another node, w/o repetition



Path = edges from a node to another node, w/o repetition

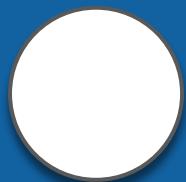


Path = edges from a node to another node, w/o repetition



Path = edges from a node to another node, w/o repetition

Quick Recap - The Basic Building Blocks



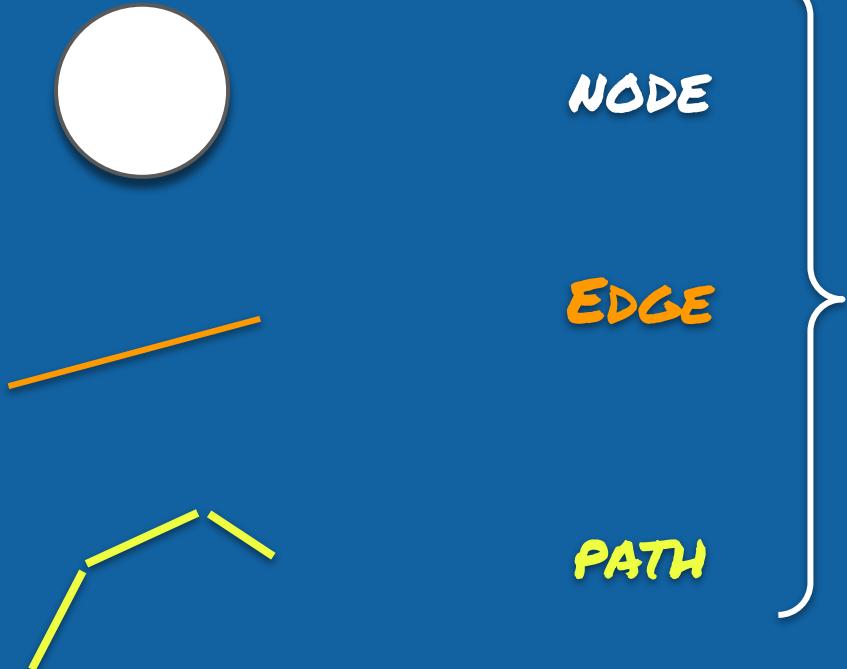
NODE

EDGE



PATH

Taking It To The Next Level



Explore more complex network structures:

- connected components
- communities

Identify nodes of interest with centrality measures:

- degree centrality
- betweenness centrality
- Pagerank

Network Analysis in Python

The best way to learn is to get our hands dirty with real data and work out an example, so...

`pip install networkx`

Network Analysis in Python

Why NetworkX?

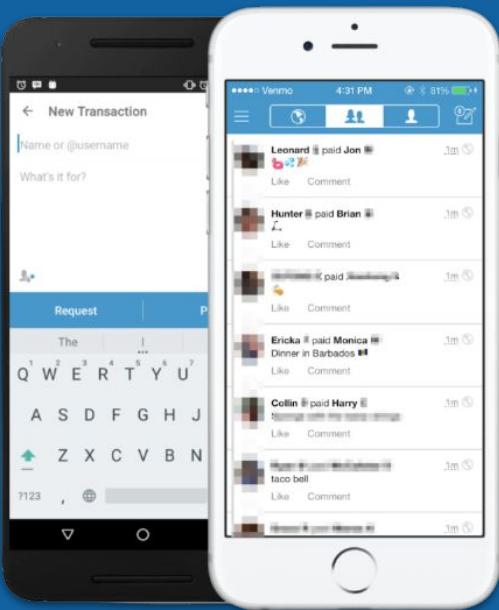
- Comprehensive (mostly)
- Friendly API
- Good documentation

Shortcomings

- Scalability to very large graphs
- Visualization of very large graphs

(other options: snap.py, graph-tool, iGraph and Gephi)

Example - Venmo

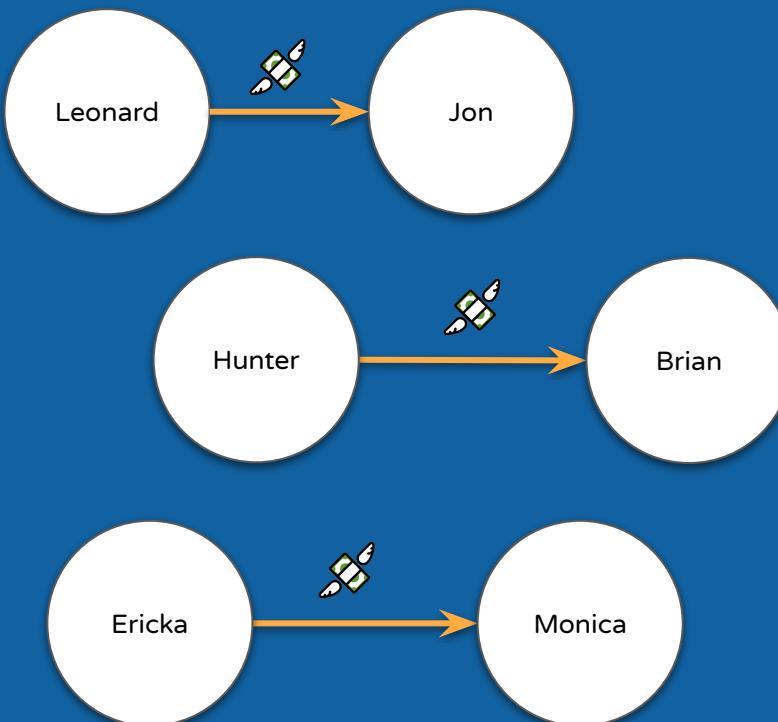


Tabular Representation (Boring)



Sender	Receiver
Leonard	Jon
Hunter	Brian
Ericka	Monica

Network Representation (Not Boring)



Motivating Examples

You are a product data scientist for a peer-2-peer payments app.

- **Customer service team:** who's customer service call should we answer first?
- **Marketing team:** who should we spend our “influencers” marketing budget on?
- **Product team:** who should we recruit as a beta tester? Who are our power users?
- **User research team:** who should we interview for product dev\usability studies?
- **Biz dev team:** who uses the app to run a business? (homework assignment)

Make Python Run Now

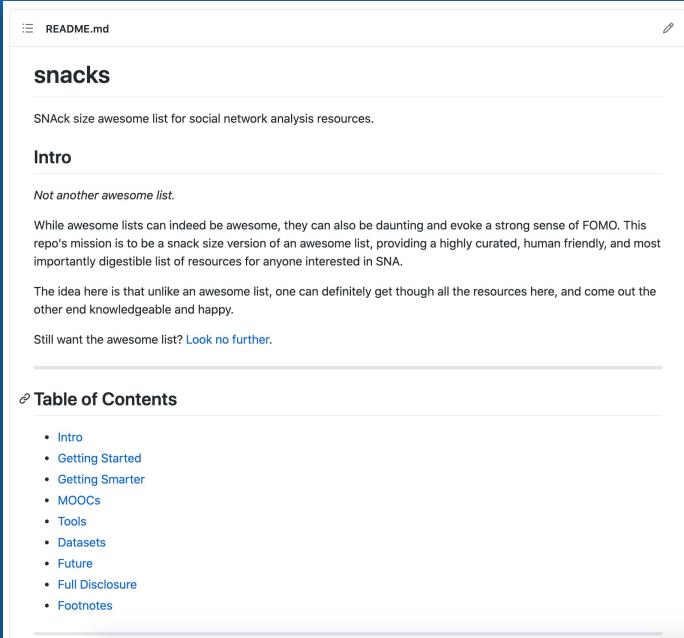
Closing Remarks

Conclusion

- We answered real world business questions using network analysis in 25 minutes.
- Network analysis is a powerful, widely applicable tool.
- We barely scratched the surface.
- NetworkX is a great place to start.
- <3 Python

Would You Like to Know More?

<https://github.com/alonnir/snacks>



The screenshot shows the content of the README.md file for a GitHub repository named "snacks". The file is titled "snacks" and describes it as a "SNACK size awesome list for social network analysis resources". The "Intro" section explains that while awesome lists can be daunting and evoke FOMO, this repo's mission is to provide a highly curated, human-friendly, and digestible list of resources for anyone interested in SNA. It emphasizes that unlike a full awesome list, one can get through all the resources here and come out the other end knowledgeable and happy. A link to "Look no further" is provided. A "Table of Contents" section at the bottom lists the following sections: Intro, Getting Started, Getting Smarter, MOOCs, Tools, Datasets, Future, Full Disclosure, and Footnotes.

```
README.md

snacks

SNACK size awesome list for social network analysis resources.

Intro

Not another awesome list.

While awesome lists can indeed be awesome, they can also be daunting and evoke a strong sense of FOMO. This repo's mission is to be a snack size version of an awesome list, providing a highly curated, human friendly, and most importantly digestible list of resources for anyone interested in SNA.

The idea here is that unlike an awesome list, one can definitely get through all the resources here, and come out the other end knowledgeable and happy.

Still want the awesome list? Look no further.

Table of Contents

- [Intro]
- [Getting Started]
- [Getting Smarter]
- [MOOCs]
- [Tools]
- [Datasets]
- [Future]
- [Full Disclosure]
- [Footnotes]
```

Take Home Challenges

- "Wikipedia surfing" challenge
 - Choose 2 pages on Wikipedia and find all the paths between them. Which nodes along the way have a high betweenness centrality?
- How would you answer the Biz Dev question with peer-2-peer payments data?
- You work for a social network and want to limit the spread of misinformation. How would you use network analysis to do that effectively?

Thank You!

Let's connect:
@alonnir