

# **DATA SCIENCE**

## **11 WEEK PART TIME COURSE**

**Week 10 – Network Analysis**  
**Monday 23rd May 2016**

1. Tasks from Wednesday
2. Network Analysis
3. Lab
4. Discussion

---

**DATA SCIENCE PART TIME COURSE**

---

# **WHAT IS NETWORK ANALYSIS?**



Many types of real-world problems involve dependencies between observations.

For example:

- Town planners are looking at vehicular flows through a city
- Sociologist want to understand how people influence others that they know (if at all)
- Biologists want to know how proteins regulate the actions of other proteins
- Information agencies want to discover groups of adversaries

Problems involving dependencies can often be modelled as graphs, and scientists have developed methods for answering these questions called network analysis.

# WHAT IS A GRAPH?

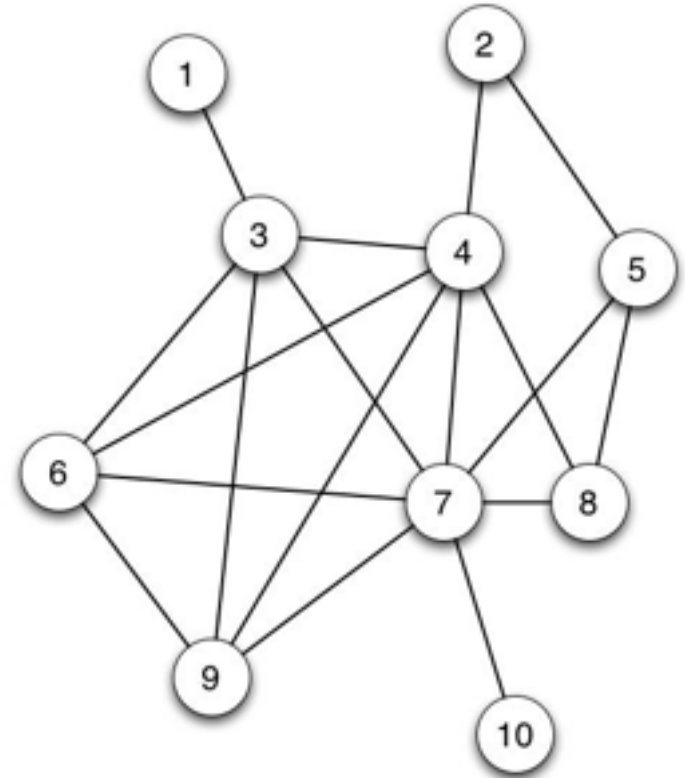
---

7

A graph consists of a nodes (or vertices) and are connected by edges.

For example the nodes may represent people and the edges are there if a friendship exists.

How many nodes and edges are there?

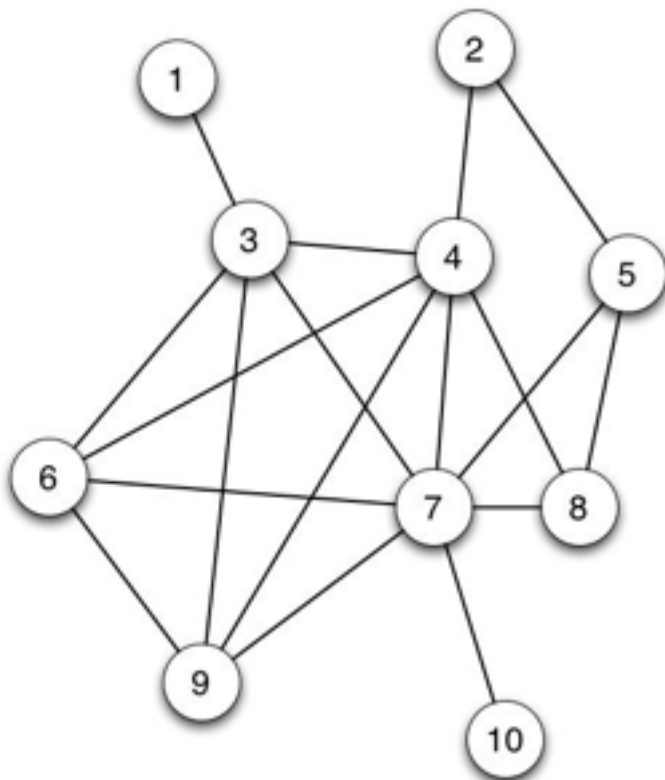


# WHAT IS A GRAPH?

8

Nodes = 10

Edges = 18

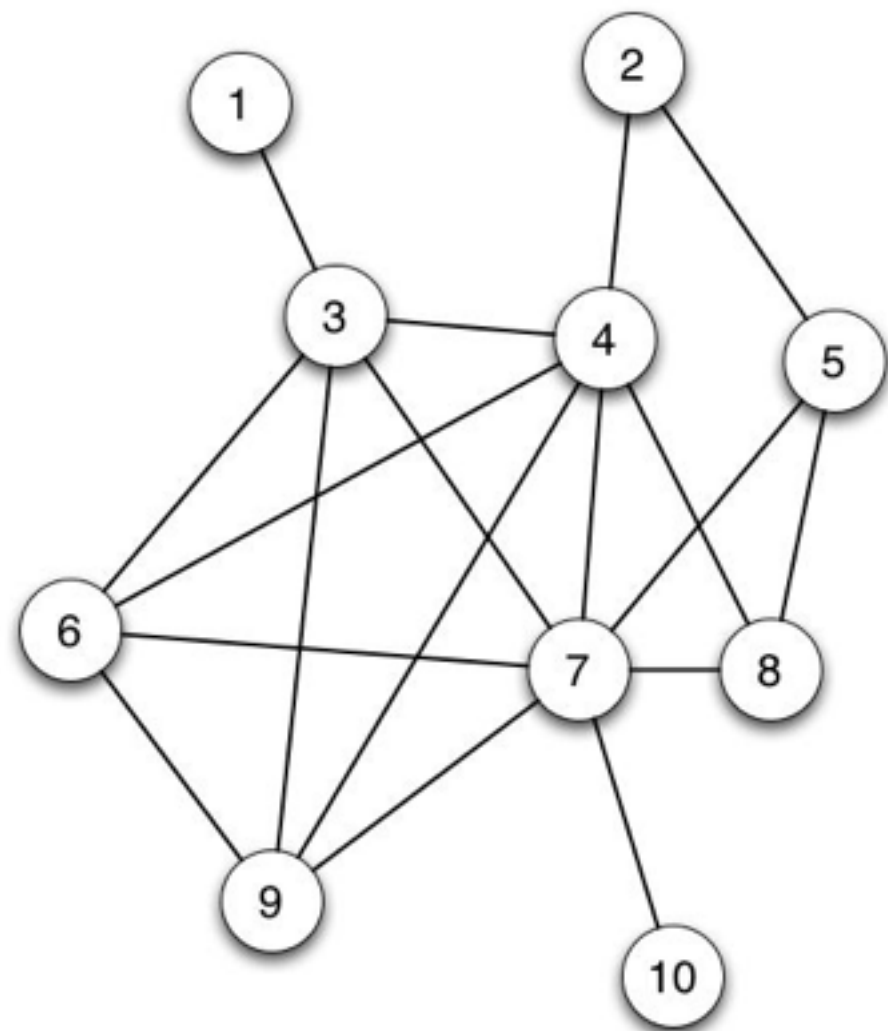




How do we know who the key actors (nodes) are in our network ?

We need to calculate some measures of importance (or centrality).

What are some of the ways you might measure an individual's centrality (there are multiple measures).



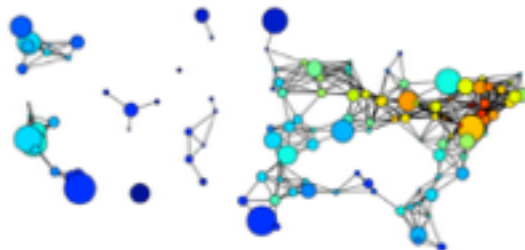
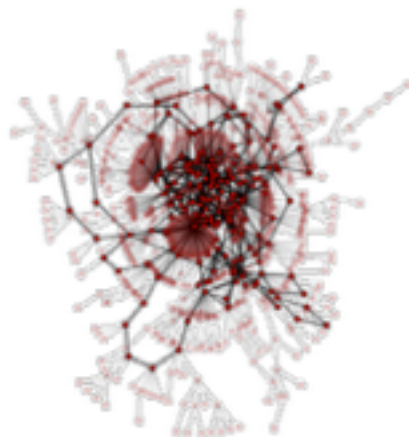
- › Degree Centrality - number of edges a node has
- › Closeness Centrality - the reciprocal of the sum of the shortest path distances from one node to all  $n-1$  other nodes. Since the sum of distances depends on the number of nodes in the graph, closeness is normalized by the sum of minimum possible distances  $n-1$ . Higher values of closeness indicate higher centrality
- › Betweenness Centrality - the sum of the fraction of all-pairs shortest paths that pass through the node  $v$
- › Eigenvector centrality - computes the centrality for a node based on the centrality of its neighbours
- › Page Rank - count the number and quality of links to a page to determine a rough estimate of how important the website is. The

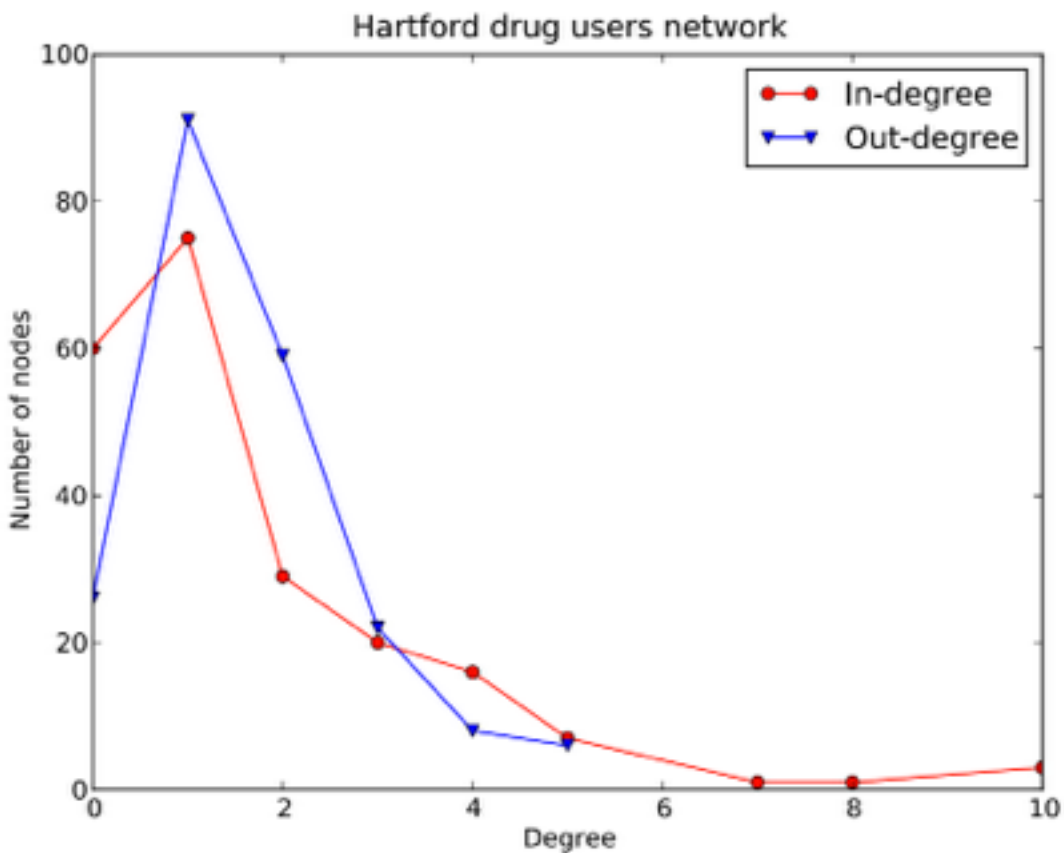
### NetworkX

- › Tool to study the structure and dynamics of social, biological, and infrastructure networks
- › Ease-of-use and rapid development
- › Open-source tool base that can easily grow in a multidisciplinary environment with non-expert users and developers
- › An easy interface to existing code bases written in C, C++, and FORTRAN
- › To painlessly slurp in relatively large nonstandard data sets

# HOW DO WE VISUALISE A GRAPH?

13





# HOW DO WE VISUALISE A GRAPH?

15



# HOW DO WE VISUALISE A GRAPH?

16



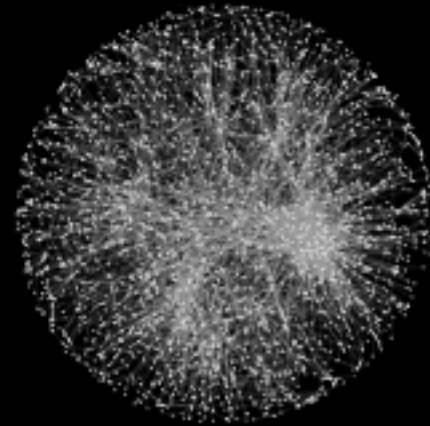
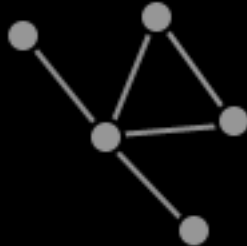


THIS NETWORK

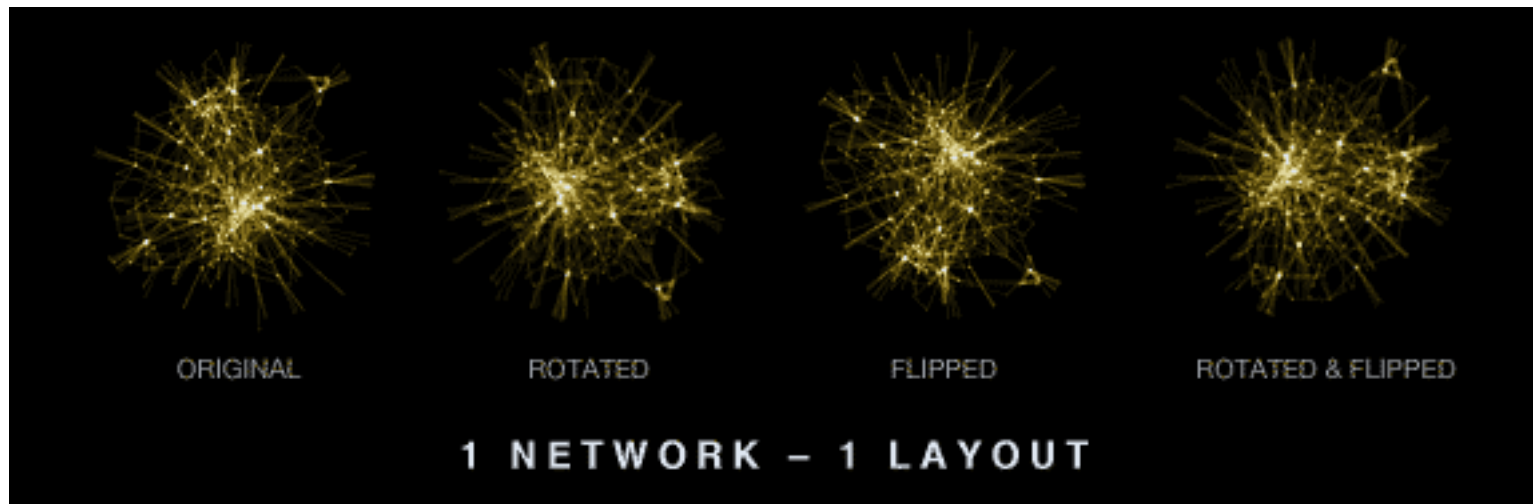
DOESN'T NEED

SOPHISTICATED

VISUALIZATION

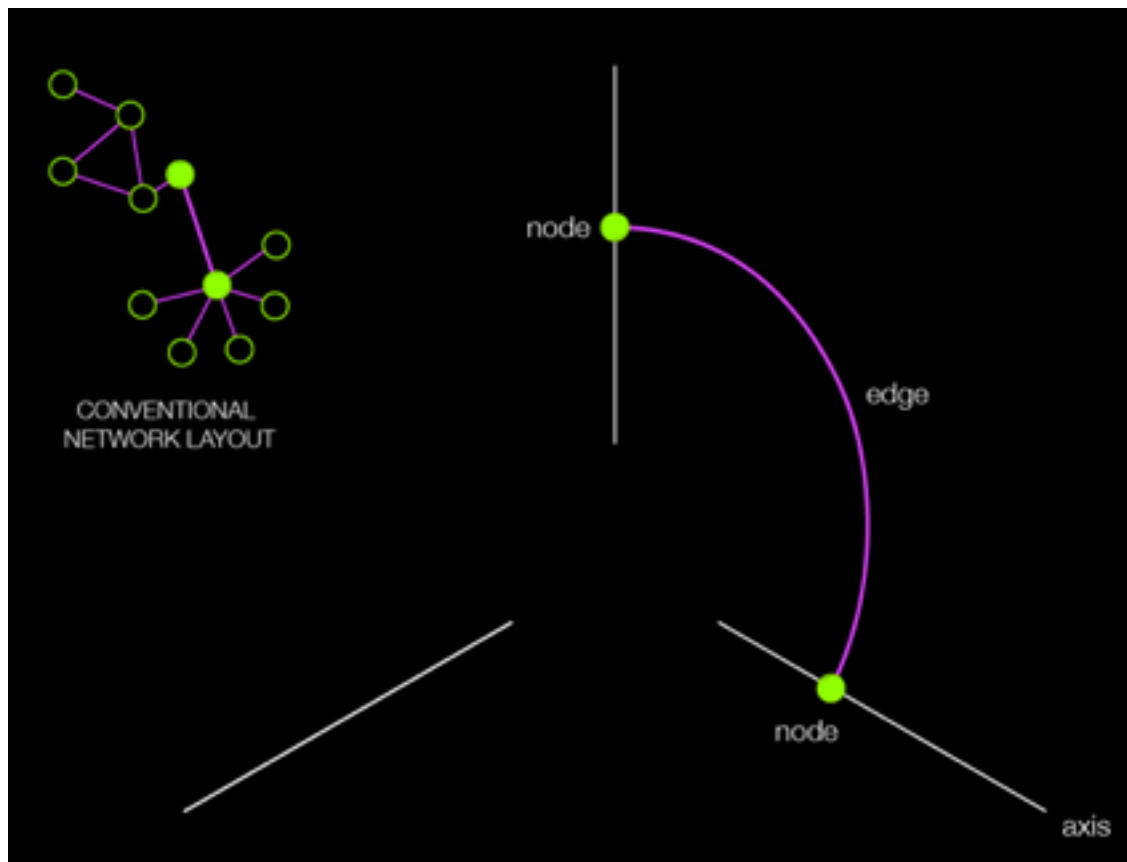


THIS ONE DOES



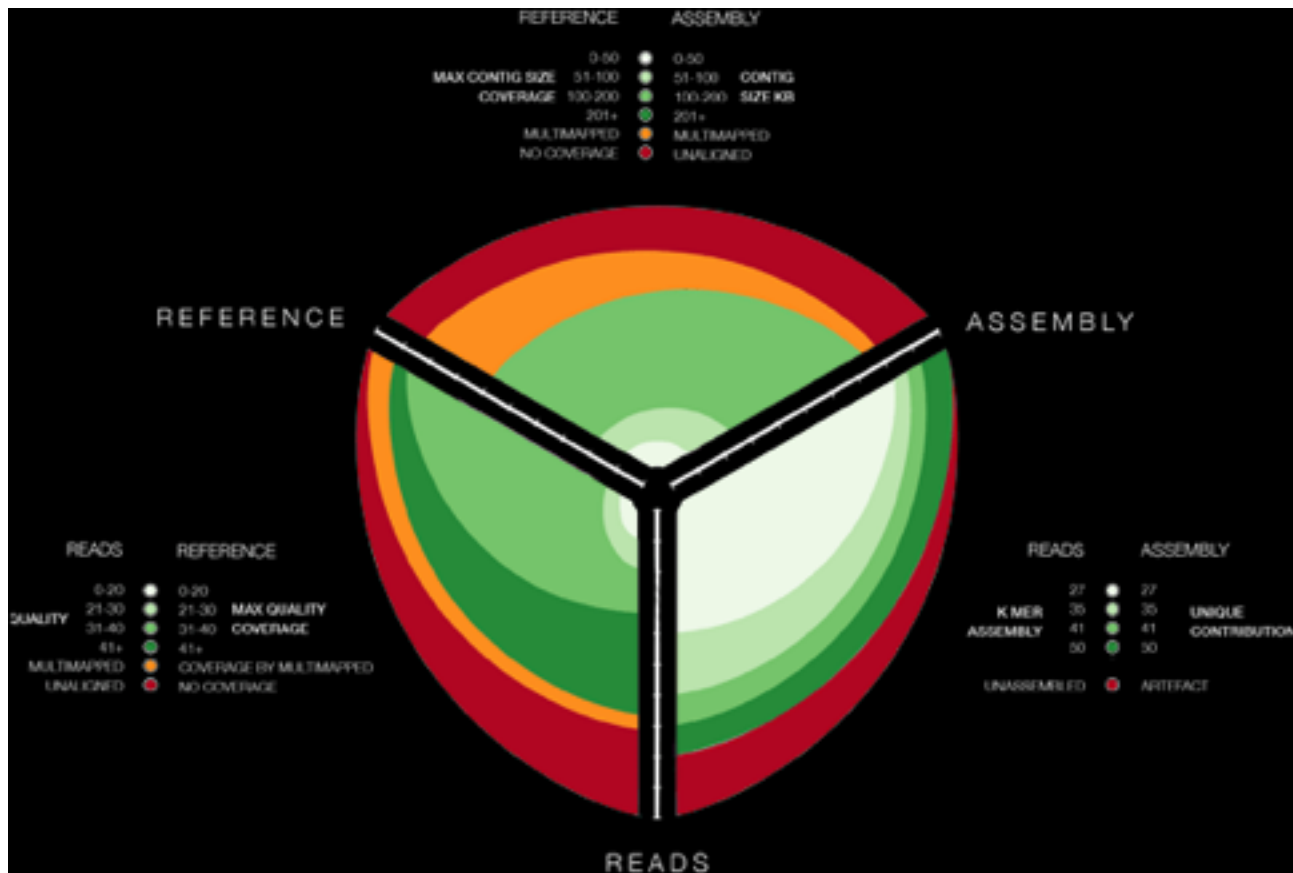
# HOW DO WE VISUALISE A GRAPH?

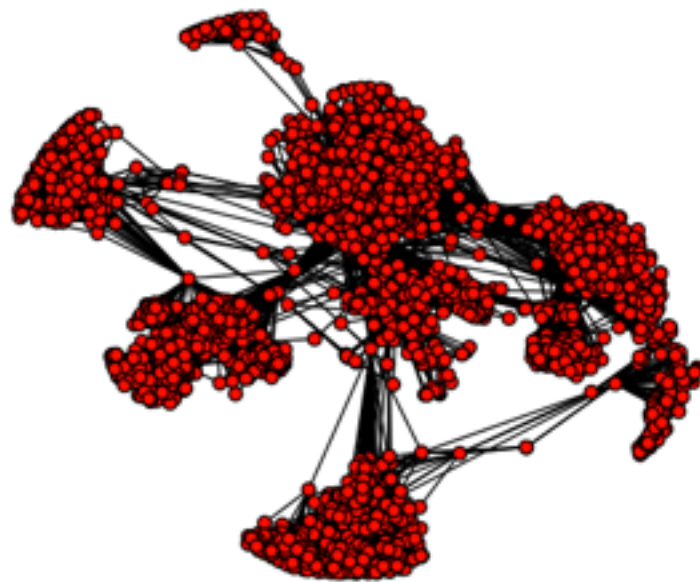
19



## HOW DO WE VISUALISE A GRAPH?

20





We all know on our Facebook page we have friends we made at different places and stages in our lives. We might have some work friends, school friends, some friends we made travelling, etc...

How might we find good groupings of communities?

The criteria for finding good communities is similar to that for finding good clusters.

We want to maximize intra-community edges while minimizing inter-community edges.

Formally, the algorithm tries to maximize the modularity of network, or the fraction of edges that fall within the community minus the expected fraction of edges if the edges were distributed by random. Good communities should have a high number of intra-community edges, so by maximizing the modularity, we detect dense communities that have a high fraction of intra-community edges.





---

**DATA SCIENCE PART TIME COURSE**

---

# **CASE STUDY BONAFIDE**

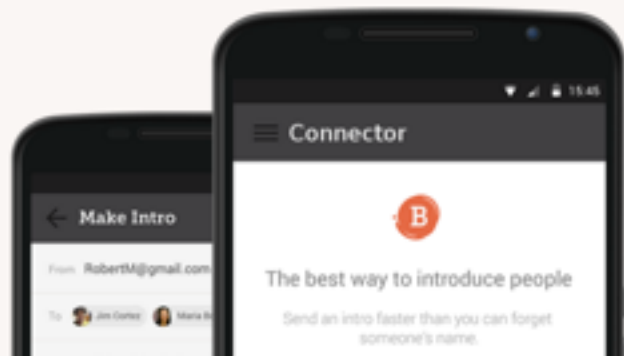


# Find your dream team.

Bonafide helps you find great people to work with  
via trusted introductions.



Sign up for Bonafide Connector



## Connector

**Make an intro in less time than it takes to forget a name.**

Connect people in seconds in your own words from your own email address. Bonafide connects with Gmail to find quotes you've written about people in previous intros, letting you write intros more quickly.

Eva Ho, Investor/Advisor/Board Member:

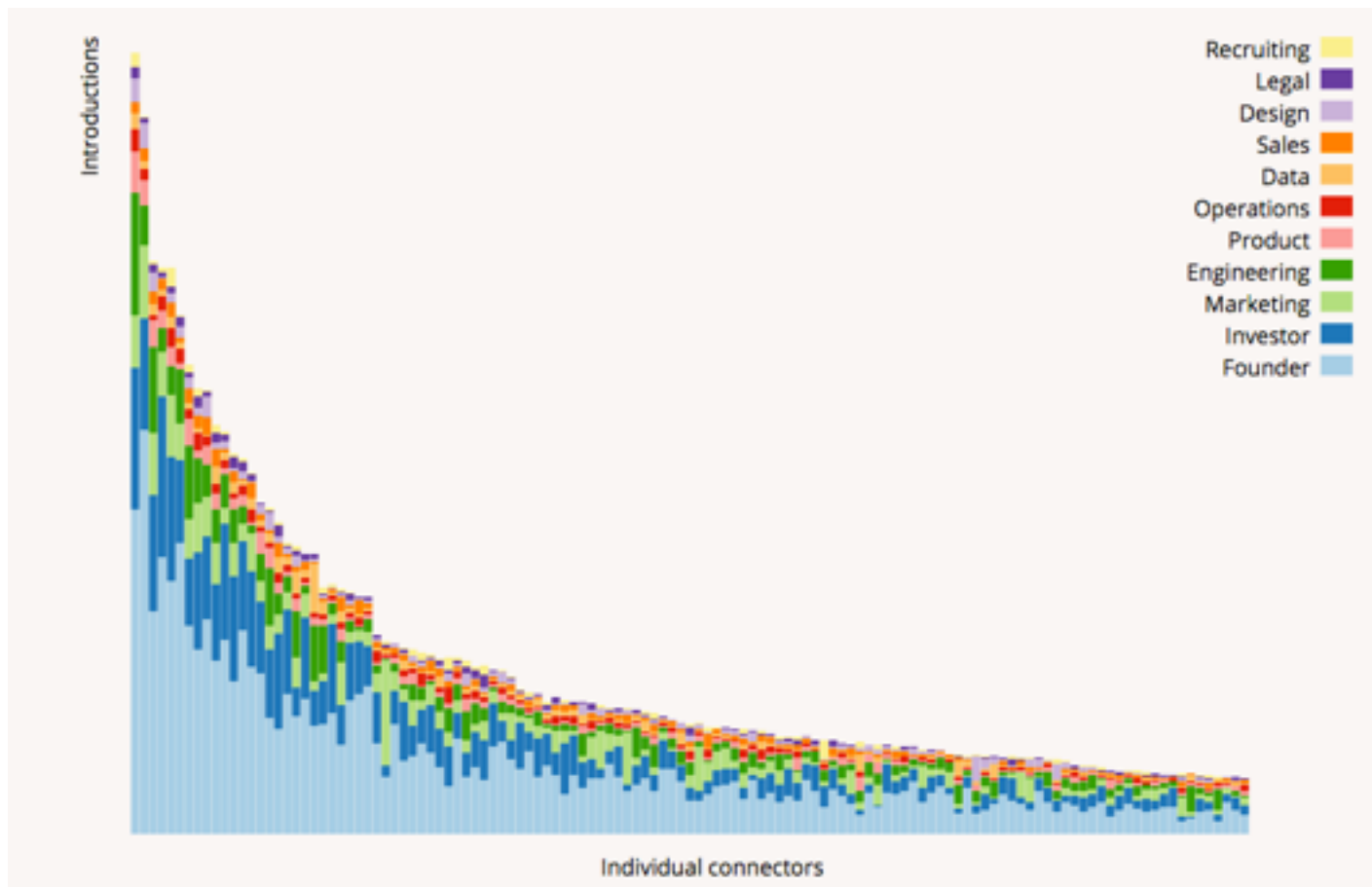


Investor

Tony Wu, Engineer at Uber:



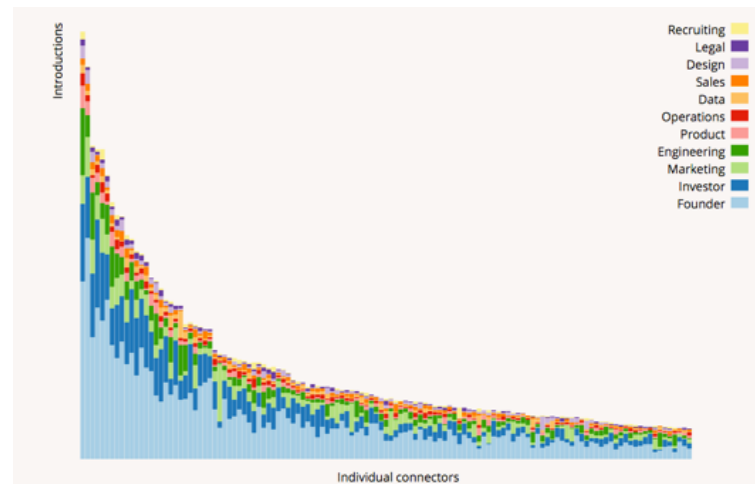
Engineering



Two things stand out from the figure:

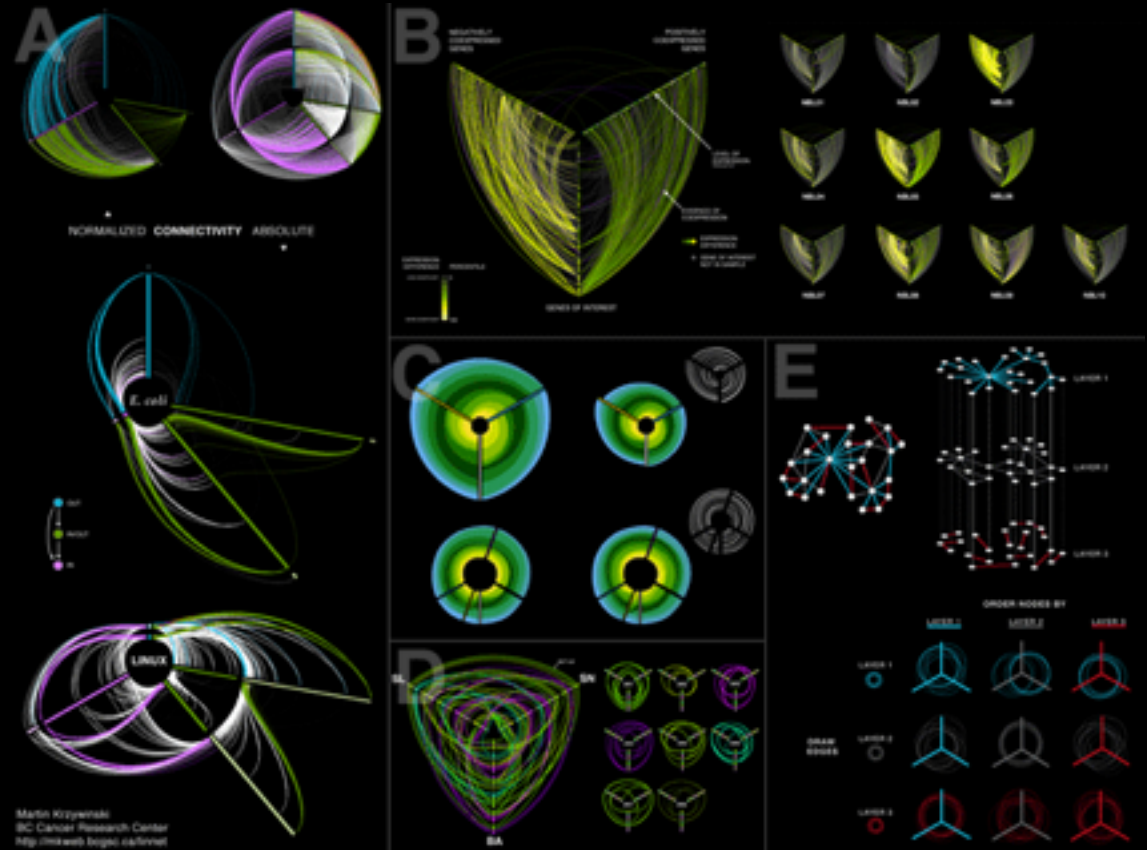
- › A lot of people introduce founders
- › Heavy connectors disproportionately introduce investors

<https://bonafide.co/blog/which-professions-top-connectors-introduce>



# DATA SCIENCE PART TIME COURSE

# LAB



---

**DATA SCIENCE - Week 10 Day 1**

---

# **REVIEW**

- **Review**
- **Tasks**
- **Review**

# **DISCUSSION TIME**

## **Course**

- **How's it going?**
- **Anything in particular during the past weeks that was challenging?**
- **How has the pace been?**
- **Anything you'd like to review next week?**



# DATA SCIENCE - Week 10 Day 1

---

## Task List

- ☐ Try running the Spark cluster (or ec2 instance)
- ☐ Read the following articles:
  - ☐ <http://www.wired.com/2016/01/googles-go-victory-is-just-a-glimpse-of-how-powerful-ai-will-be/>
  - ☐ <http://www.wired.com/2014/01/geoffrey-hinton-deep-learning>
  - ☐ <https://sites.google.com/site/deepernn/home/blog/briefsummaryofthepaneldiscussionatdlworkshopicml2015>
- ☐ Install TensorFlow

## **DATA SCIENCE - Week 10 Day 2**

---

### **Task List**

☐ **EMAIL ONE OF THE GUEST PRESENTERS OF THIS CLASS**

# Life Inc.

HOW THE WORLD BECAME A CORPORATION  
AND HOW TO TAKE IT BACK

DOUGLAS RUSHKOFF



## DATA SCIENCE - Week 10 Day 1

---

WELL DONE  
EVERYONE !!