Handong Global Univ. Seminar

INTRO TO NETWORK SCIENCE

Presenter: Hankyu Jang

Date: May 31, 2021

University of Iowa comp|epi computational epidemiology research

The University of Iowa

Bio

EDUCATION

University of Iowa, Iowa City, IA — 2018-2023
Ph.D. in Computer Science

Indiana University, Bloomington, IN — 2016-2018
M.S. in Data Science

Handong Global University, Pohang, Korea — 2009-2016
B.S. in Computer Science & Management, Cum Laude

WORK EXPERIENCE

Machine Learning and Data Science Intern
American Family Insurance, Madison, WI, USA

Graduate Research Assistant
Dept. of Computer Science, University of Iowa, Iowa City, IA, USA
Advisor: Dr. Alberto Segre and Dr. Sriram Pemmaraju
- Designed deep learning framework to learn patient embedding
- Designed models to detect asymptomatic infections of HAI in hospital
- Developed agent-based disease simulators for HAIs COVID-19 simulator

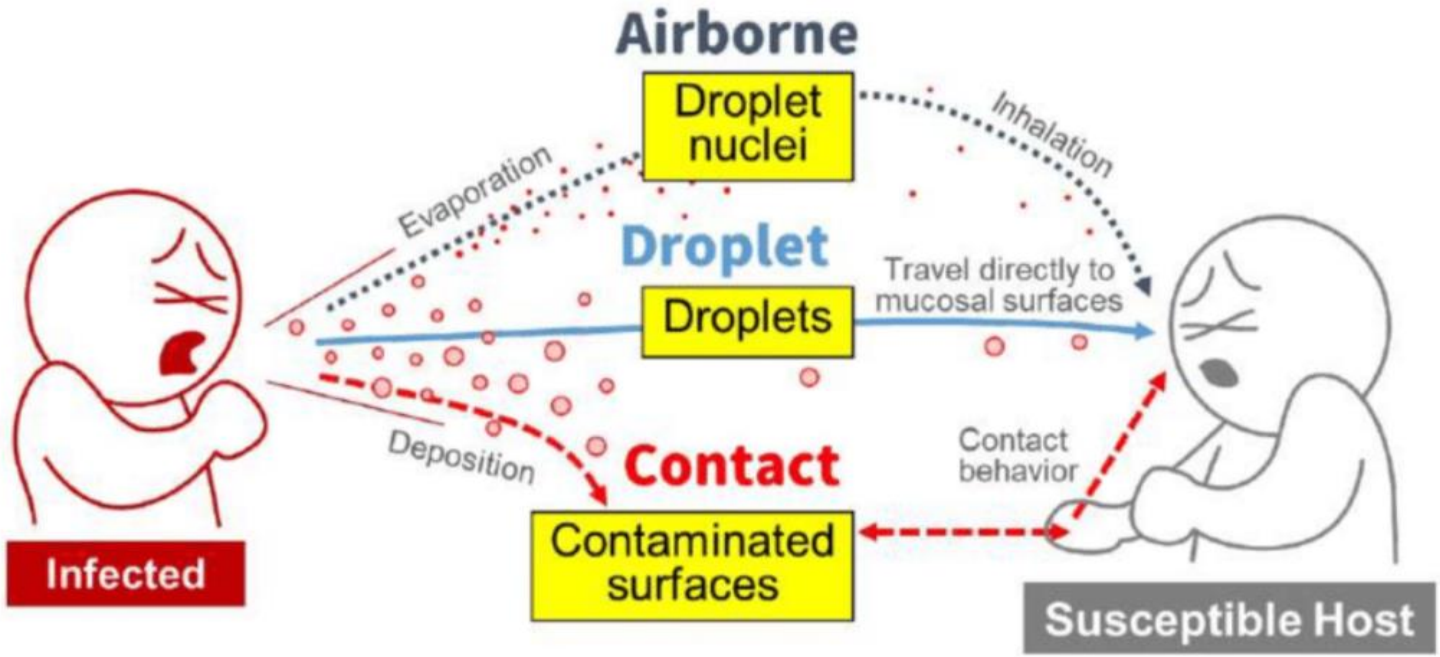
Contents

Enter your subheadline here

- Intro
- Part 1: Network science in general
  - Complex system
  - Network science basics
  - Applications: centrality, link prediction, node classification
- Part 2: Application to problems in healthcare
  - How to design interventions to reduce the spread of COVID-19?
  - How to capture medical history of patients?
- Q & A

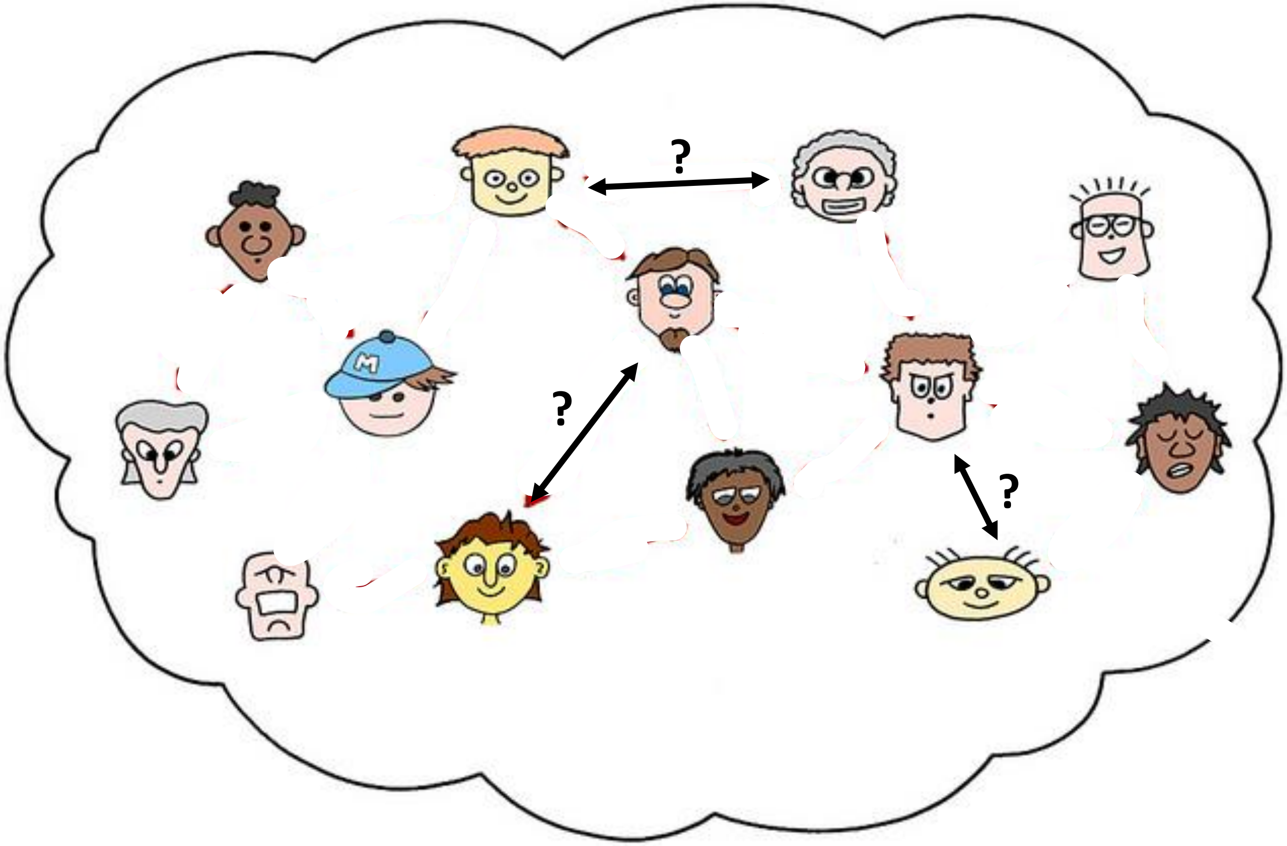
Part1 Network science in general

Complex system (disease)



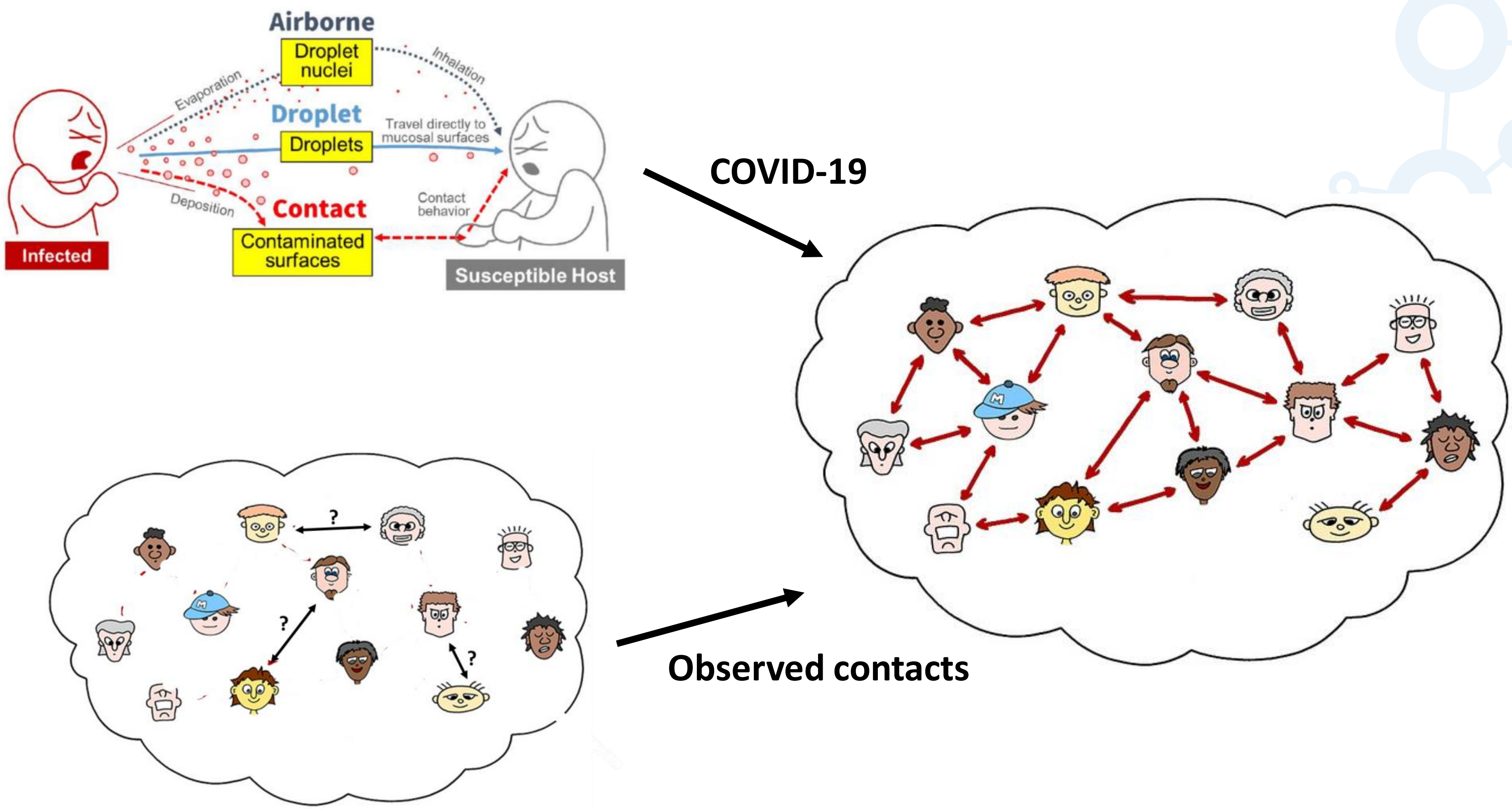
Gameiro da Silva, M. An analysis of the transmission modes of COVID-19 in light of the concepts of Indoor Air Quality. Doi: 10.13140. 2020

Complex system (contacts)



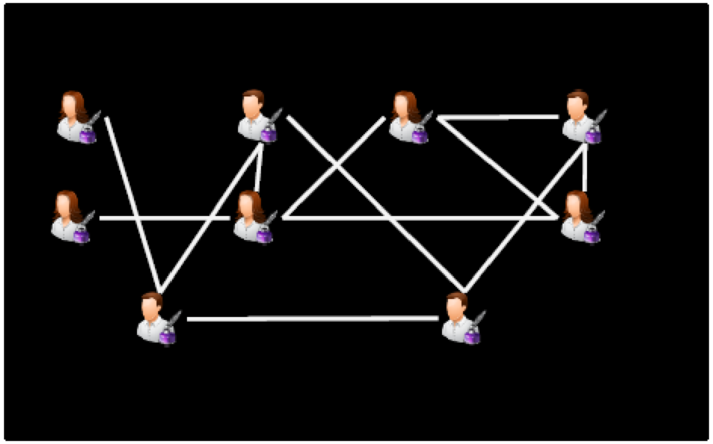
https://www.straby.com/how-to-build-a-contact-network.html

How to represent complex system?



What is a network?

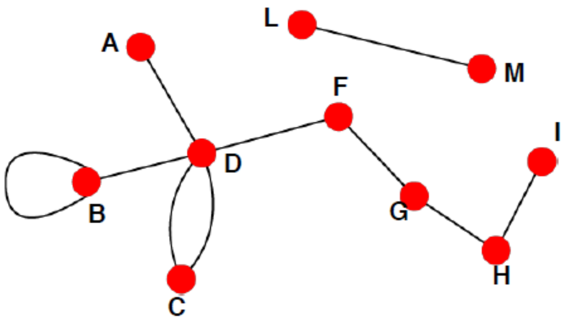
- A network is a collections of nodes with relations between some nodes



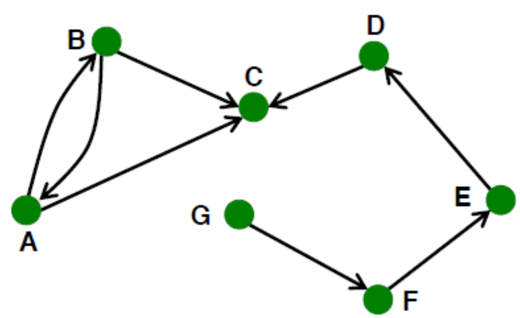
Object: nodes, vertices *N*
Relations: links, edges *E*
System: graphs, networks *G*(*N*, *E*)

Undirected vs Directed

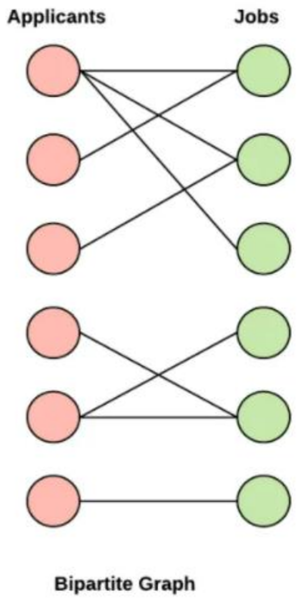
- Undirected
  - Links are symmetrical
  - Examples
    - Friendships (on FB!)
    - Collaborators



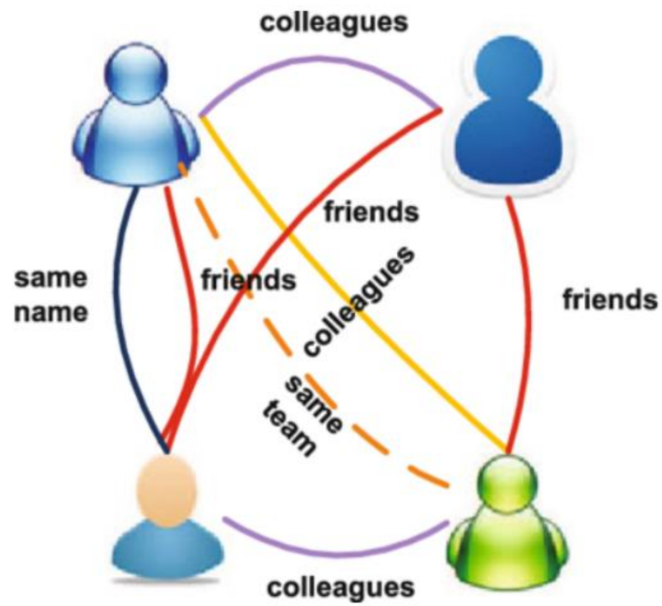
- Directed
  - Links are directed
  - Examples
    - Following on Twitter
    - Phone calls



Bipartite, multi relational



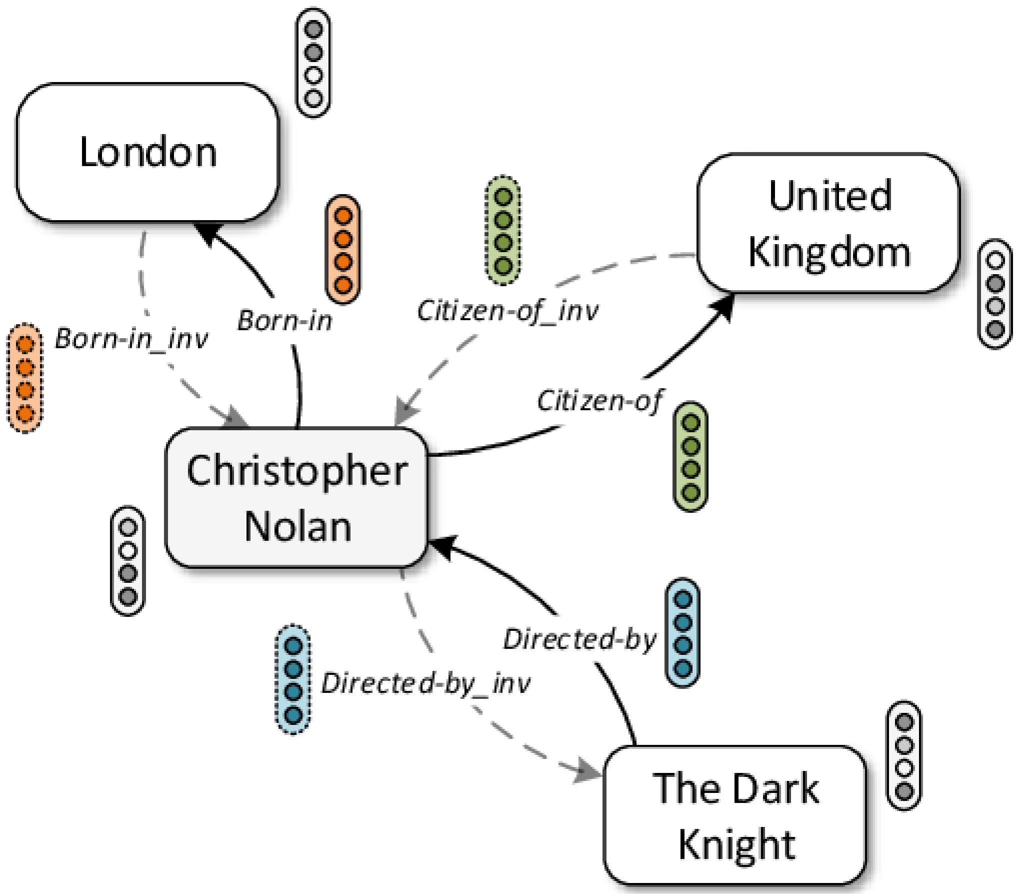
Node type (2)
- Job applicant
- Job



Edge type (>=2)
- Friend
- Colleague
- Same team

Wu, Zhiang et al. (2015). Discovering Communities in Multi-relational Networks

Heterogeneous



Vashishth, Shikhar et al. "Composition-based Multi-Relational Graph Convolutional Networks." *ArXiv* (2020)

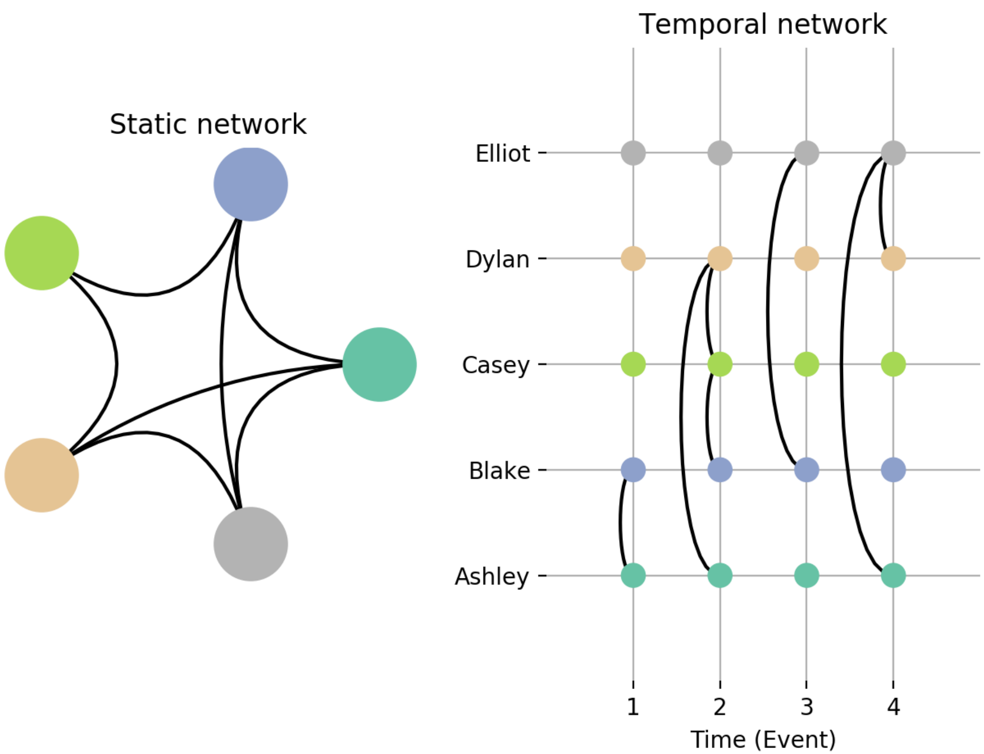
Node type

- Movie director
- Country
- City
- Movie

Edge type

- Born in
- Citizen of
- Directed by

Temporal network



Network changes over time

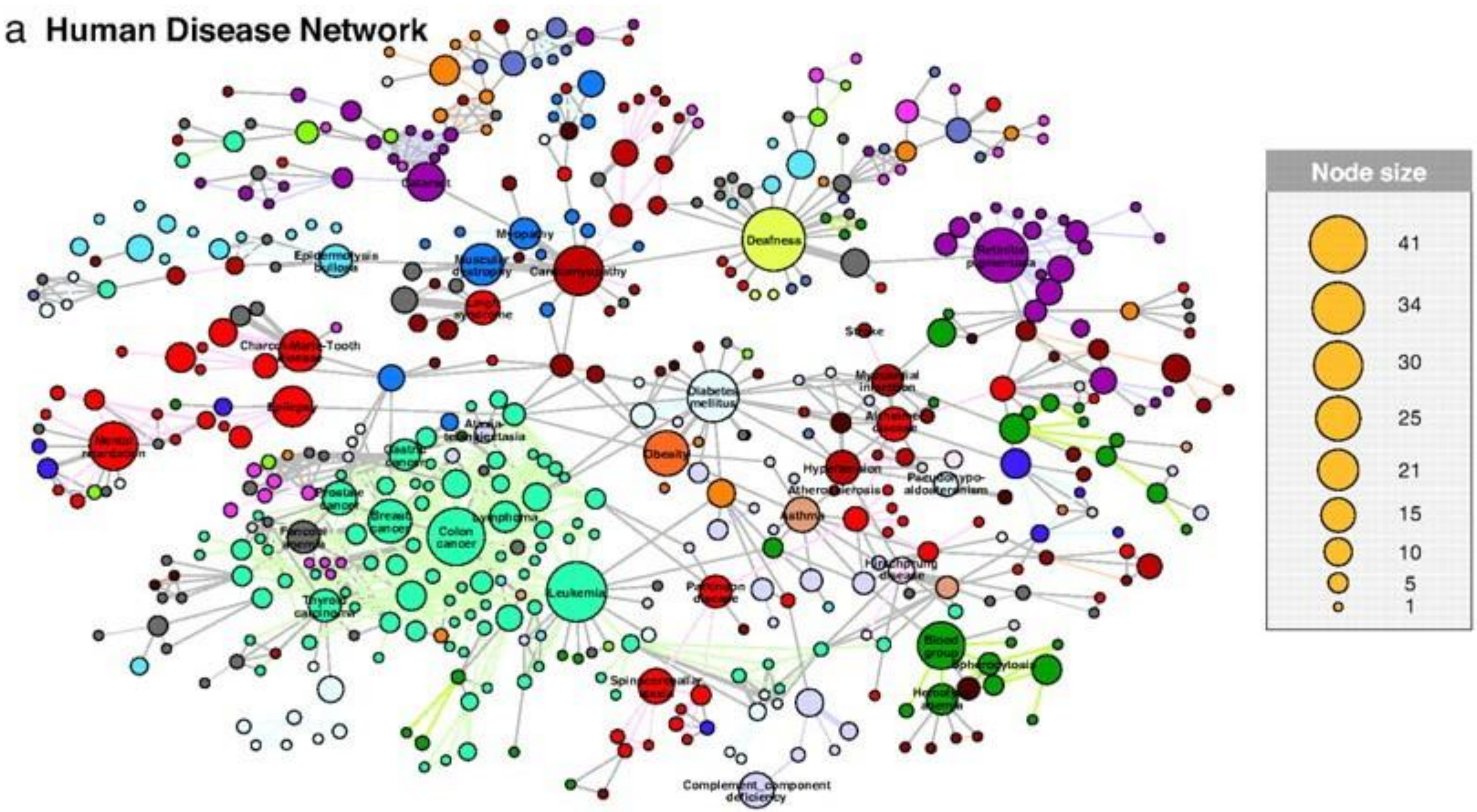
https://teneto.readthedocs.io/en/latest
/what_is_tnt.html

What are some examples of real-world networks?

Social network (Facebook)

Node: user (> 2.7 Billion)
Edge: friendship

Human disease network

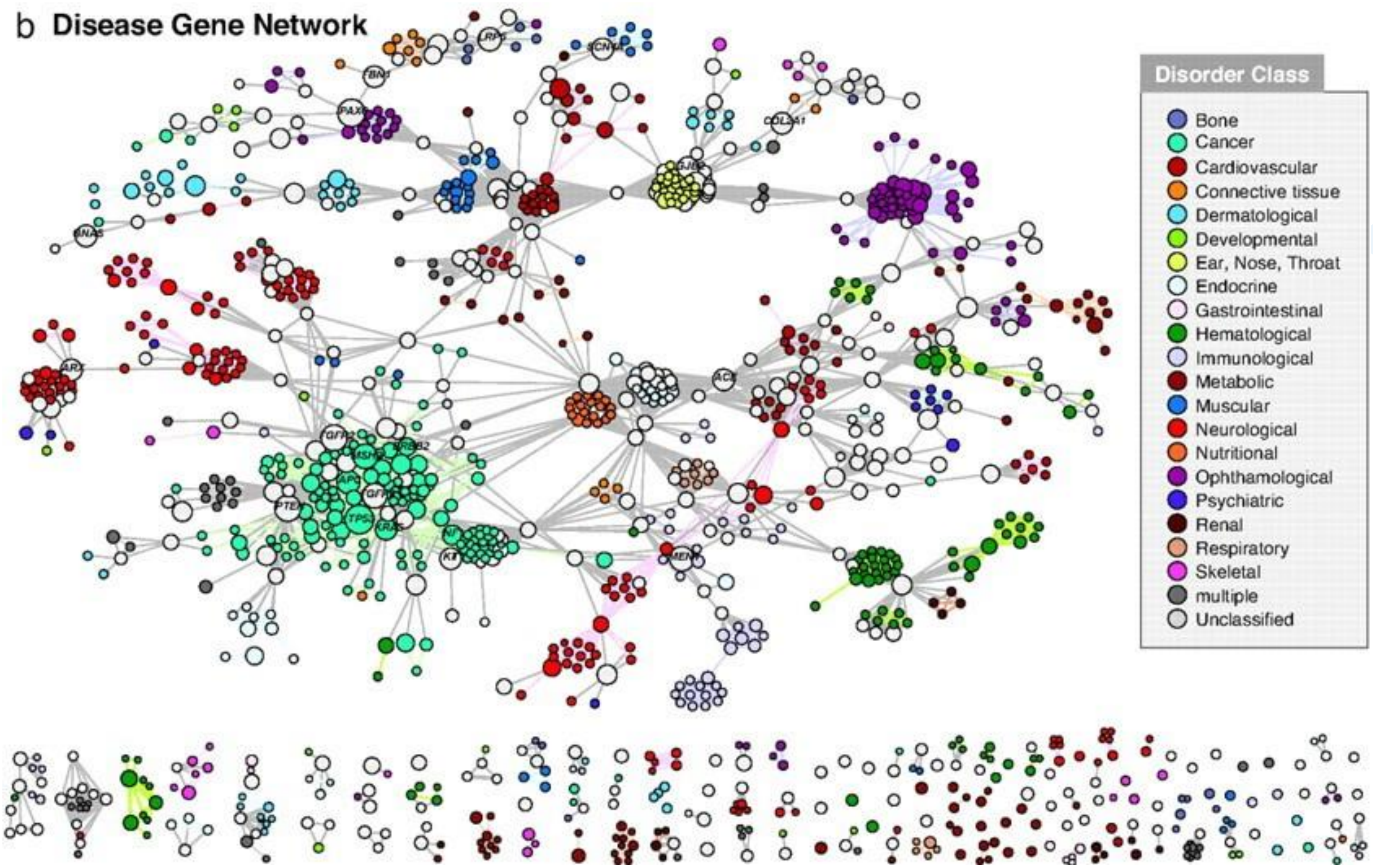


Node: disease

Edge: share genes

Human Disease Network, Barabasi 2007

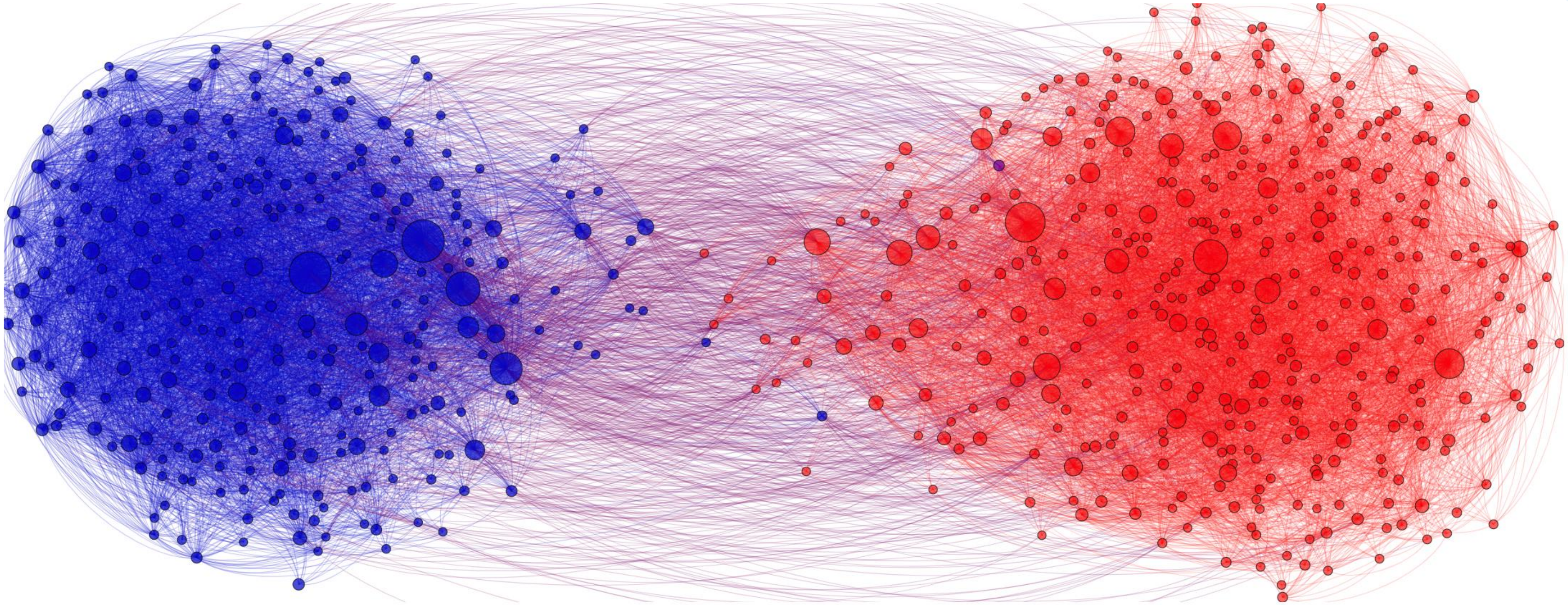
Disease gene network



Node: gene
Edge: cause same disorder

Human Disease Network, Barabasi 2007

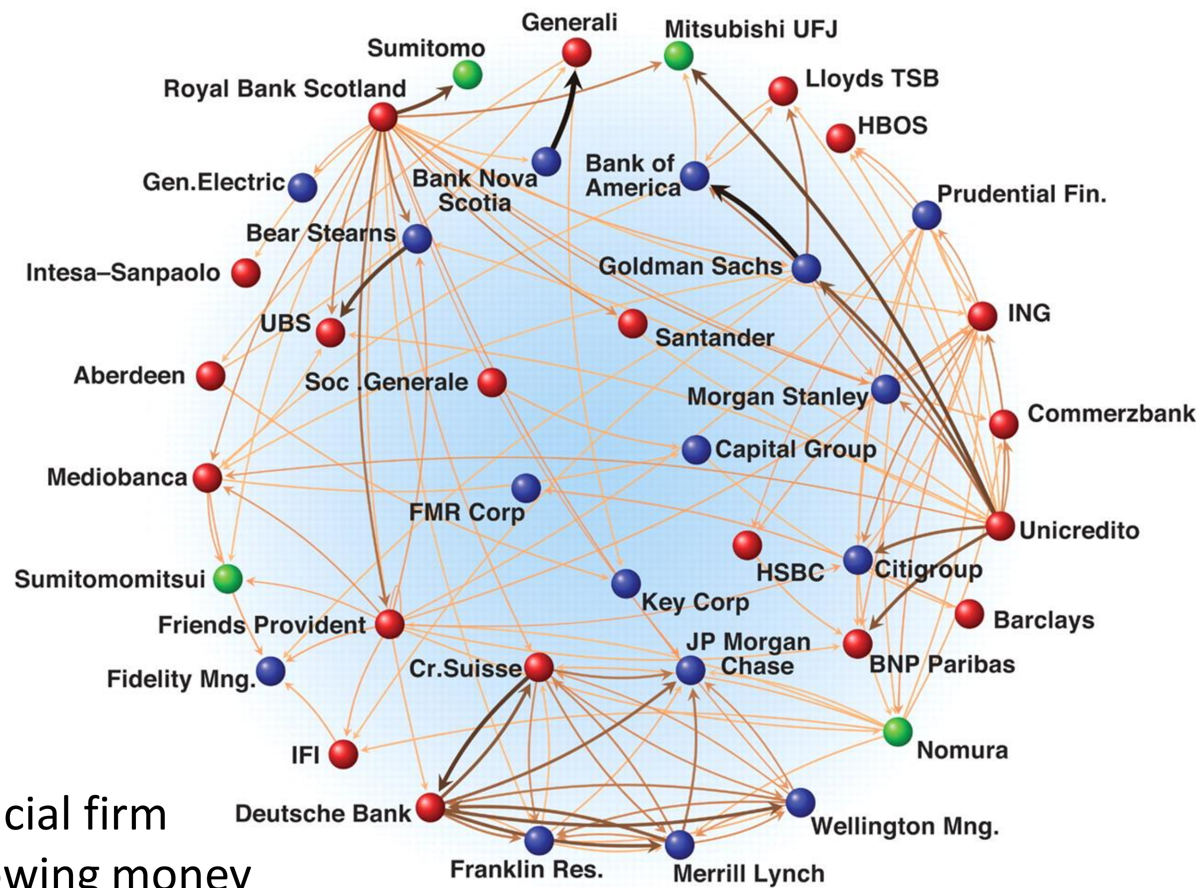
Political blog network



Node: blog
Edge: hyperlink

allthingsgraphed.com

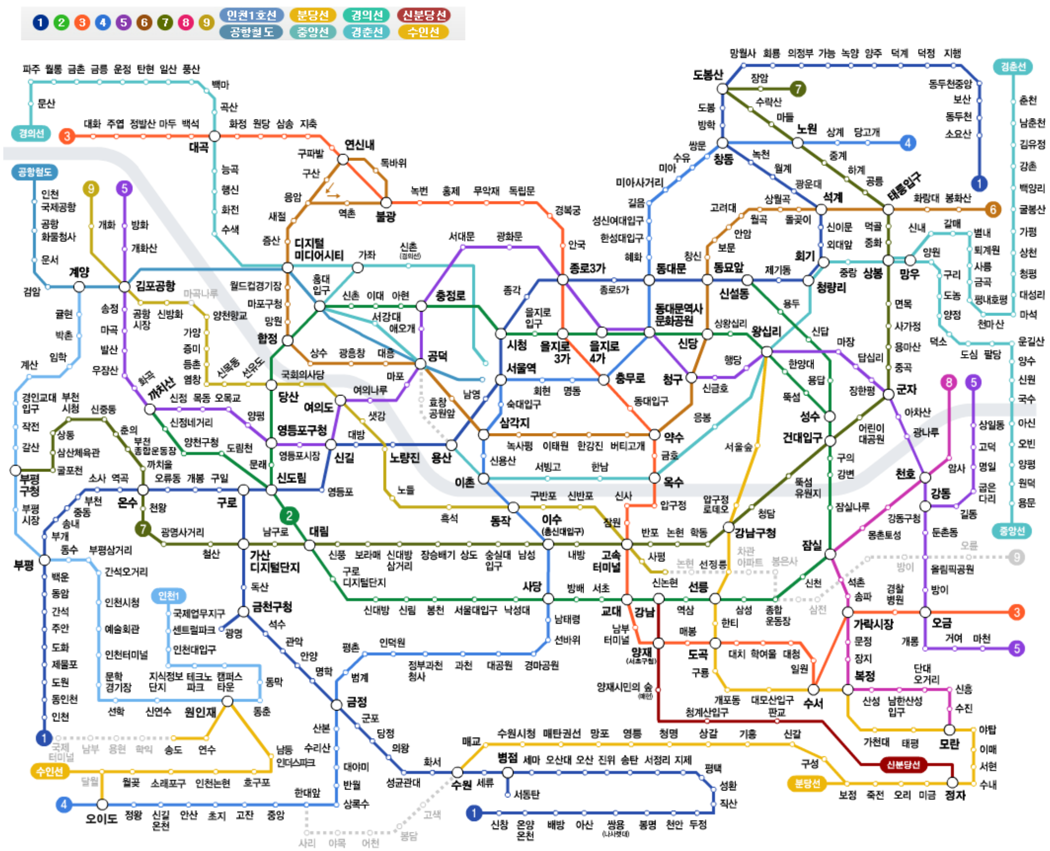
Fianancial network



Node: financial firm
Edge: borrowing money

Schweitzer, et al., Economic Networks: The New Challenges. *Science 2009*

Transportation network

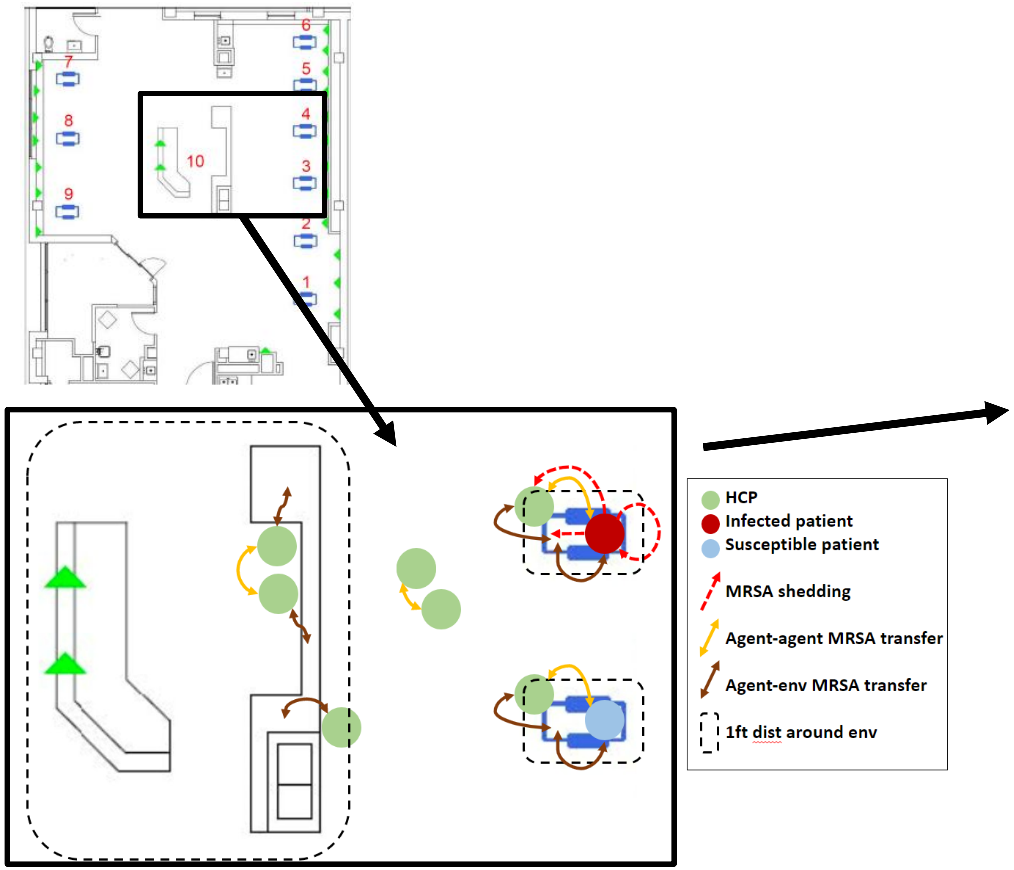


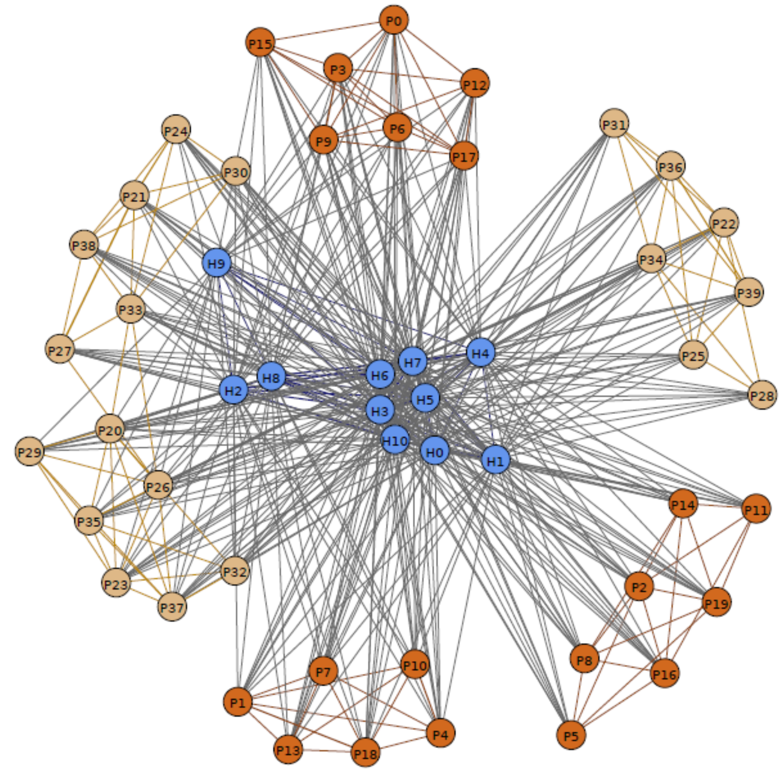
Node: station

Edge: connectivity

https://www.sisul.or.kr/open_content/skydome/introduce/pop_subway.jsp

Contact network





Node: Individual
Edge: Physical proximity

H. Jang, et al., "Evaluating Architectural Changes to Alter Pathogen Dynamics in a Dialysis Unit," **ASONAM 2019 [Best Paper Award]**

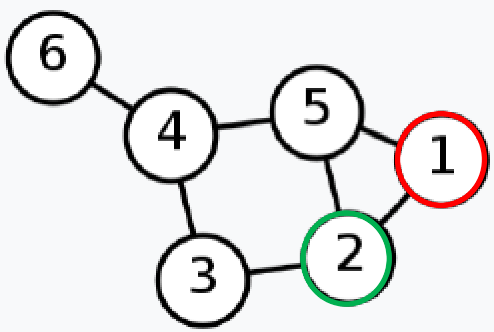
California patient transfer network

Node: hospital
Edge: patient transfer

How to represent a network?

How to represent networks?

Adjacency matrix

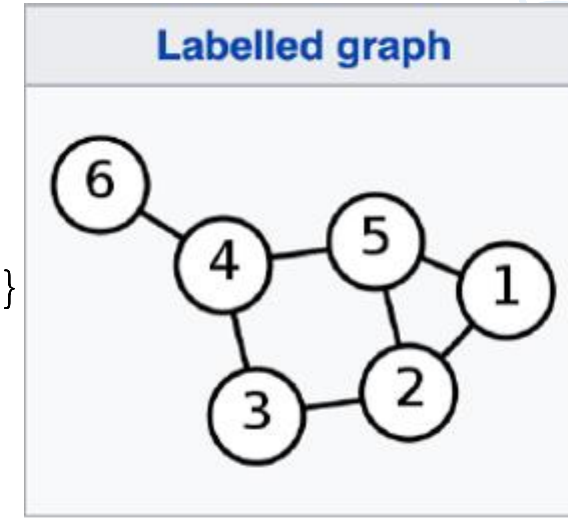
Labelled graph	Degree matrix	Adjacency matrix
	$\begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$

Edgelist
[(1, 2), (1, 5), (2, 3), (2, 5), (3, 4), (4, 5), (4, 6)]

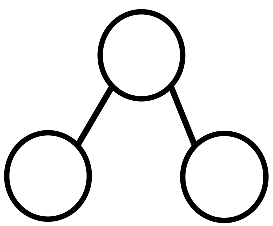
How to characterize a network?

How to characterize a network?

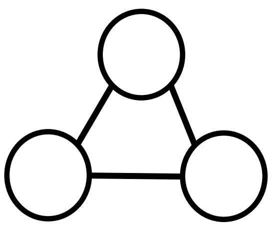
- Density: (# of edges) / (# potential edges)
- Clustering coefficient: (# of triangles)/(# 2-stars)
- Degree distribution
  - Degree count of nodes {1: *2*, 2: *3*, 3: *2*, 4: *3*, 5: *3*, 6: *1*}
- Connected: If every pair of node is *reachable*
- Diameter: largest *geodesic* distance



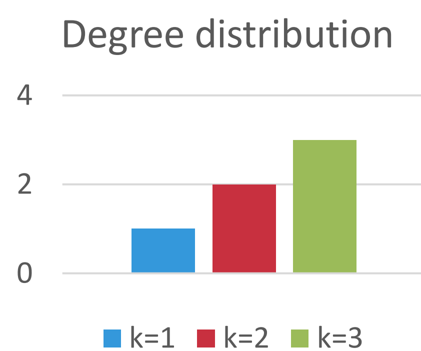
Network statistic
Density: 7 / C(6, 2)
Clustering coefficient: 3 / 11
Connected: yes
Diameter: 3



2-star

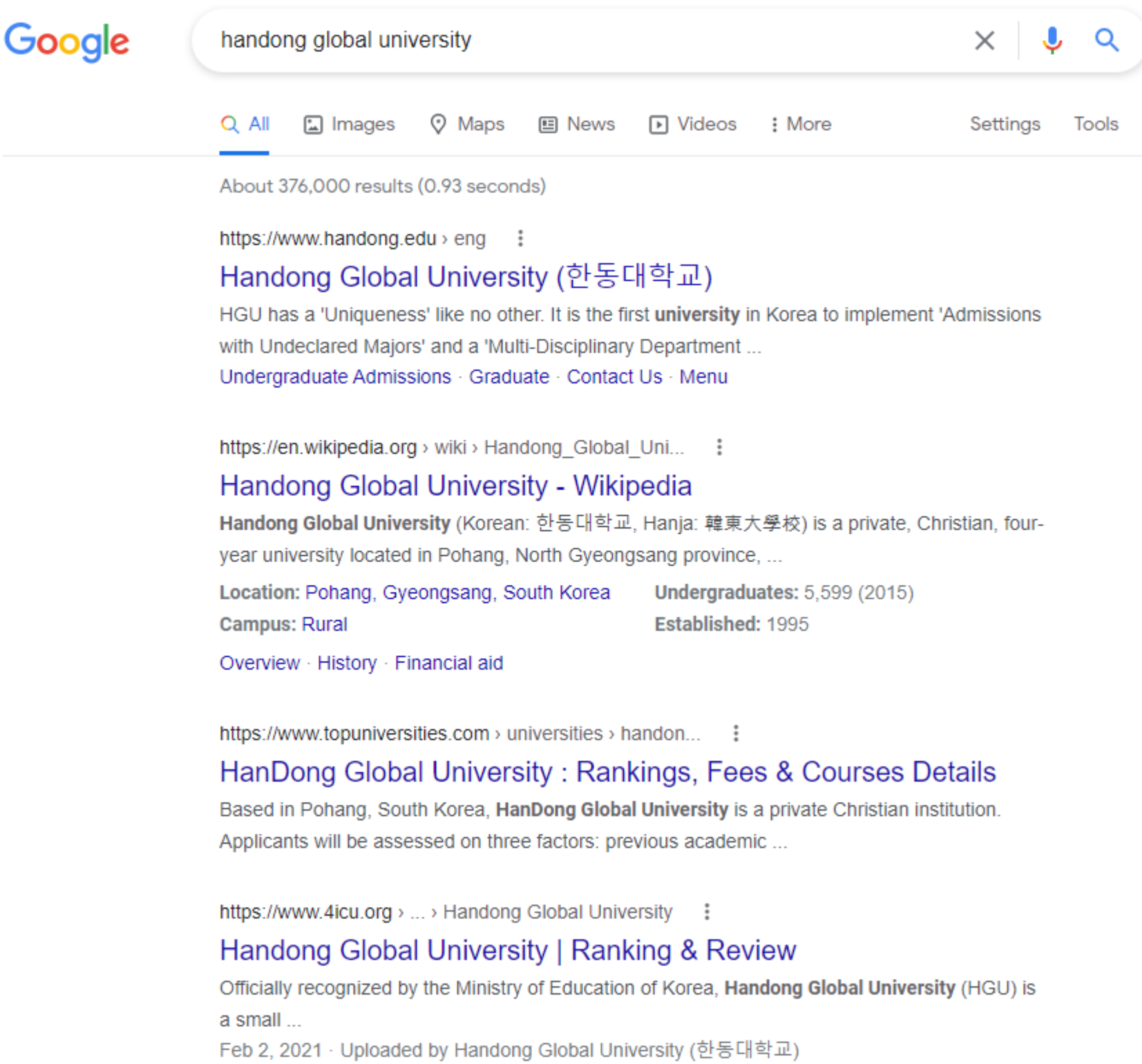


triangle



What are some applications of networks?

Application 1: Website ranking (find central node)


Google

handdong global university

All · Images · Maps · News · Videos · More · Settings · Tools

About 376,000 results (0.93 seconds)

https://www.handong.edu › eng

Handong Global University (한동대학교)

HGU has a 'Uniqueness' like no other. It is the first **university** in Korea to implement 'Admissions with Undeclared Majors' and a 'Multi-Disciplinary Department ...

Undergraduate Admissions · Graduate · Contact Us · Menu

https://en.wikipedia.org › wiki › Handong_Global_Uni...

Handong Global University - Wikipedia

Handong Global University (Korean: 한동대학교, Hanja: 韓東大學校) is a private, Christian, four-year university located in Pohang, North Gyeongsang province, ...

Location: Pohang, Gyeongsang, South Korea

Campus: Rural

Undergraduates: 5,599 (2015)

Established: 1995

Overview · History · Financial aid

https://www.topuniversities.com › universities › handon...

HanDong Global University : Rankings, Fees & Courses Details

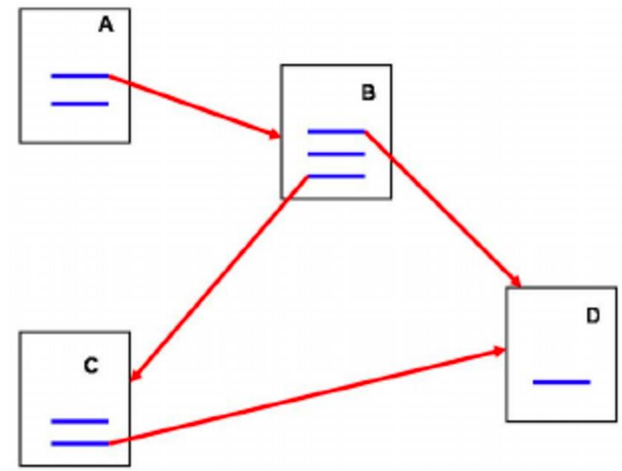
Based in Pohang, South Korea, **HanDong Global University** is a private Christian institution. Applicants will be assessed on three factors: previous academic ...

https://www.4icu.org › ... › Handong Global University

Handong Global University | Ranking & Review

Officially recognized by the Ministry of Education of Korea, **Handong Global University** (HGU) is a small ...

Feb 2, 2021 · Uploaded by Handong Global University (한동대학교)

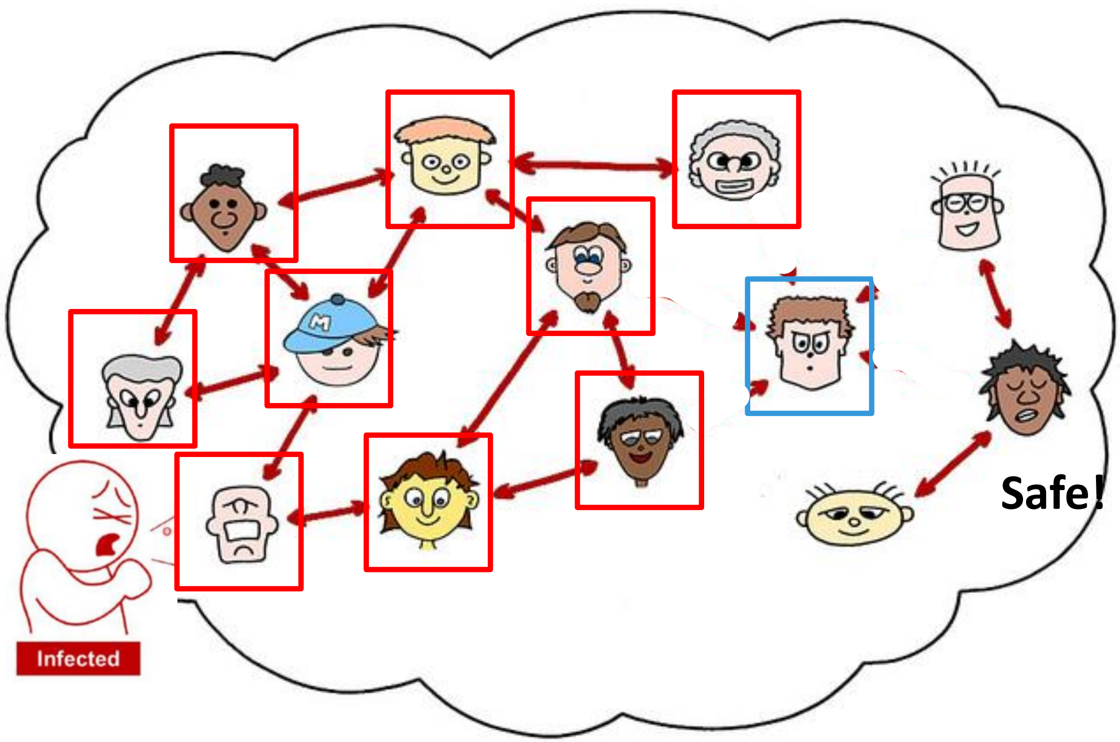


Page Rank, 1999

- Key idea: Link from page A to B is regarded as a "**vote**" for page B by A
- If many link passes through B, then B has a high ranking

Application 2: Vaccination (find central node)

- Given limited vaccine (1 dose), whom to vaccinate to ***minimize*** COVID-19 spread?

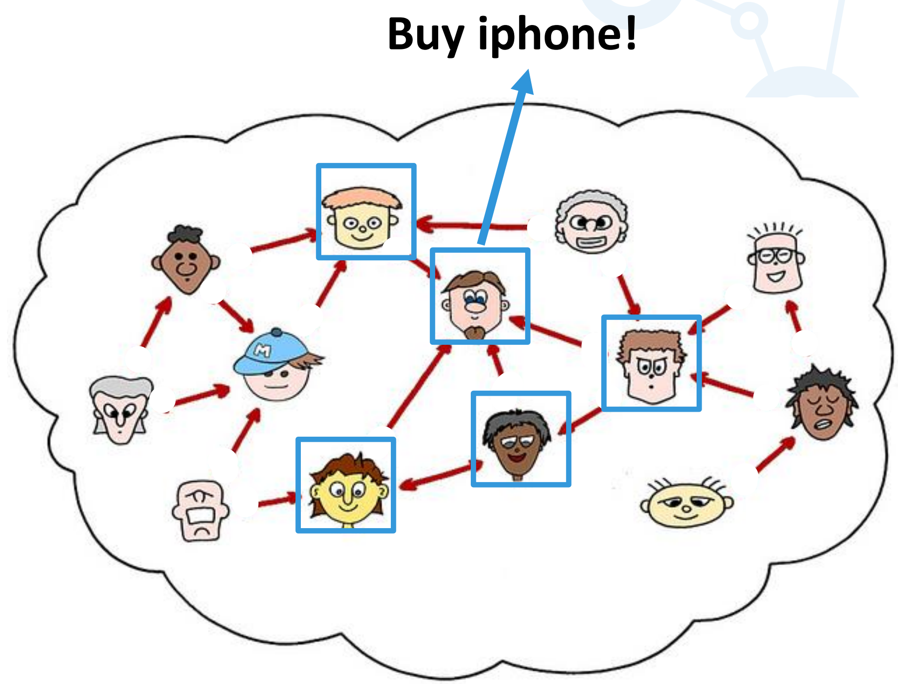


Hint: node with highest degree (***degree centrality***)

Application 3: Viral marketing (find central node)

Followers

Celeberity



Goal: ***maximize*** information spread!

Q: Apple can pay *one person* to ***advertise*** iphone 12 Pro. Whom to select?

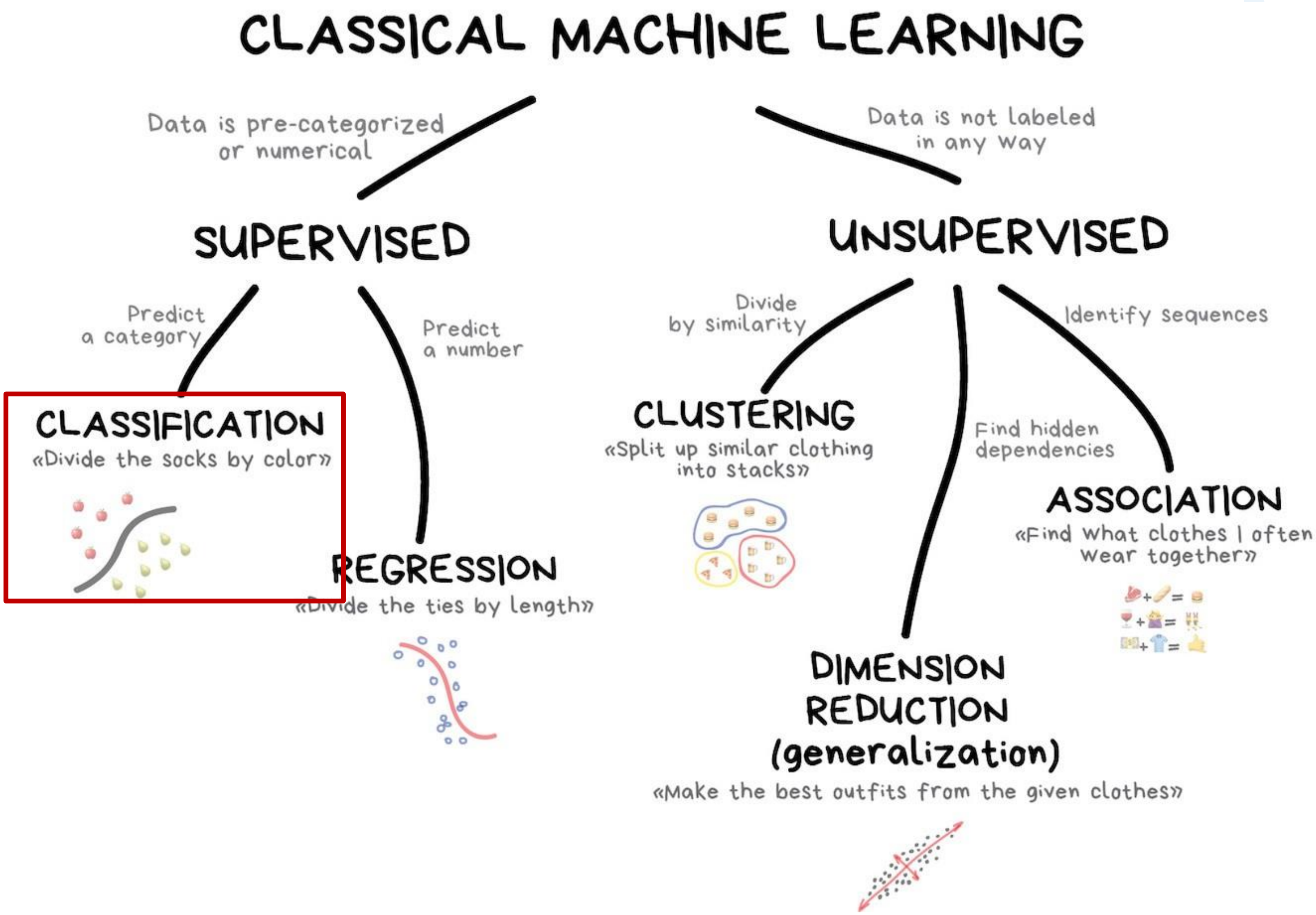
Can we use networks to solve more complex problems?

Recommender system? E.g., movie? Friend?

Patient diagnosis prediction task?

Can we use ***machine learning*** on networks for prediction tasks?

Machine learning basics

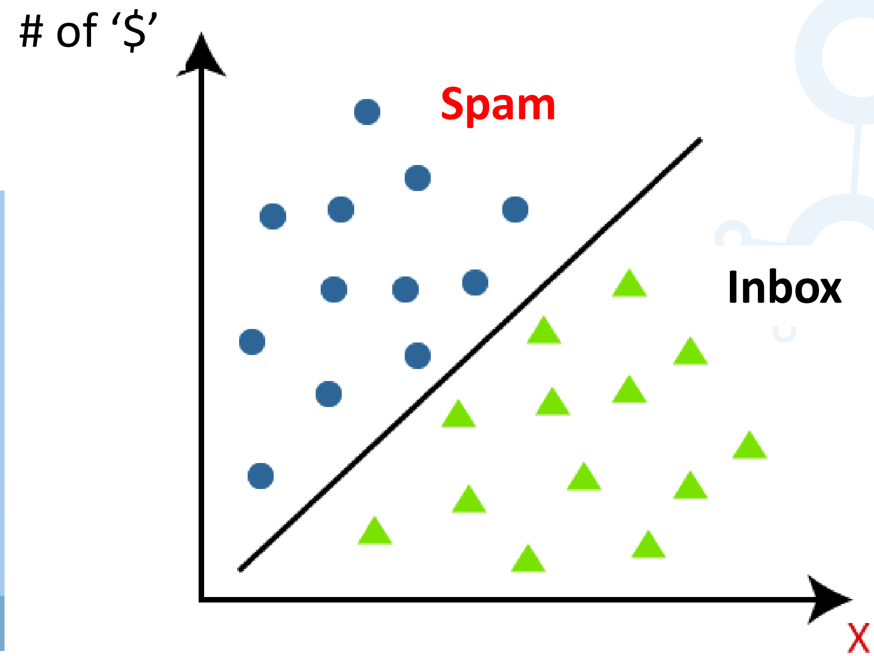


https://vas3k.com/blog/machine_learning/?fbclid=IwAR0NjjOJlZt4-KiaBGi11DskcBHAa2d6xaUchkPZdDch7pxS5sbcrZkUBJA

Classification

- Supervised learning technique to identify the category of new observations

https://www.penplusbytes.org/strategies-for-dealing-with-e-mail-spam/

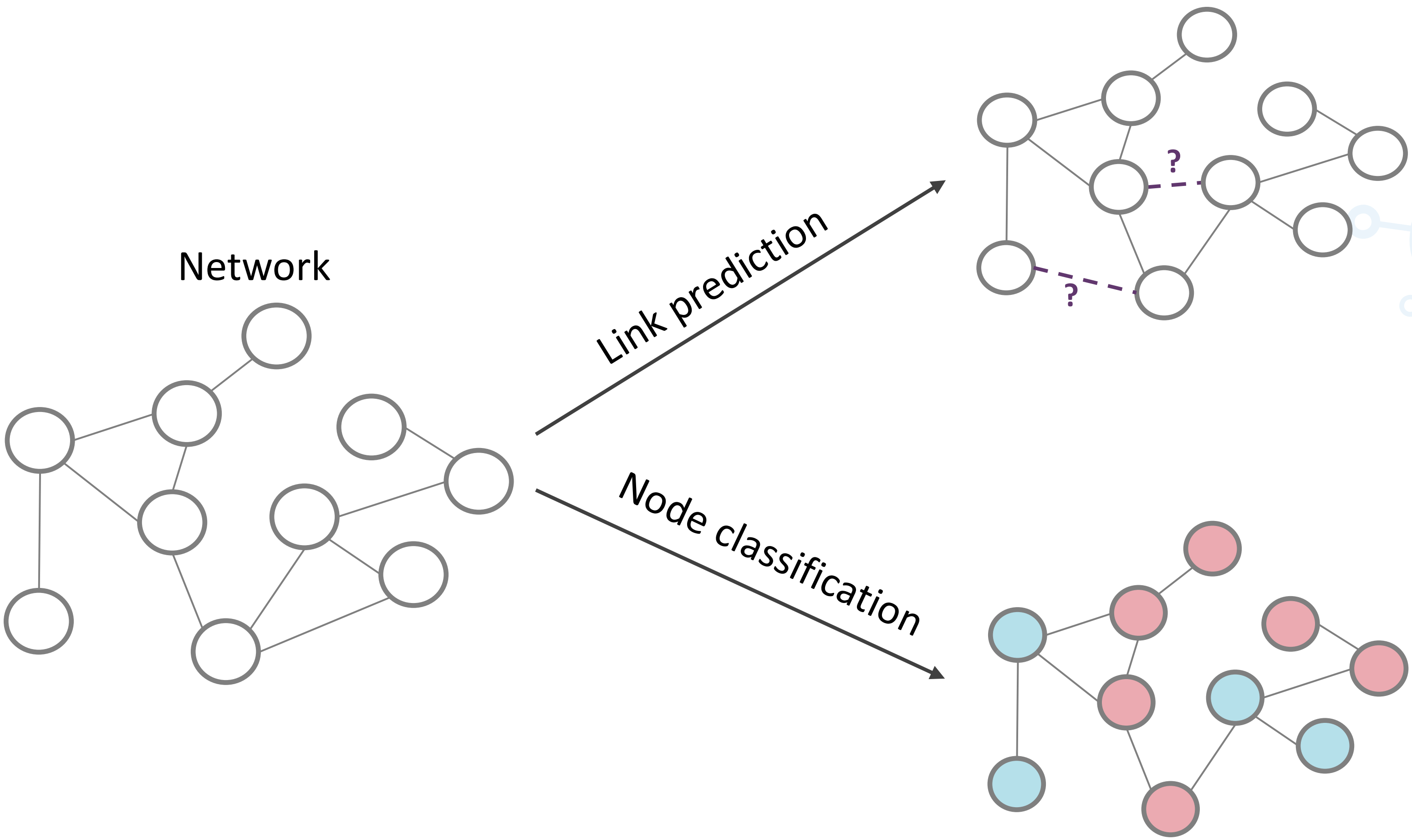


https://www.javatpoint.com/classification-algorithm-in-machine-learning

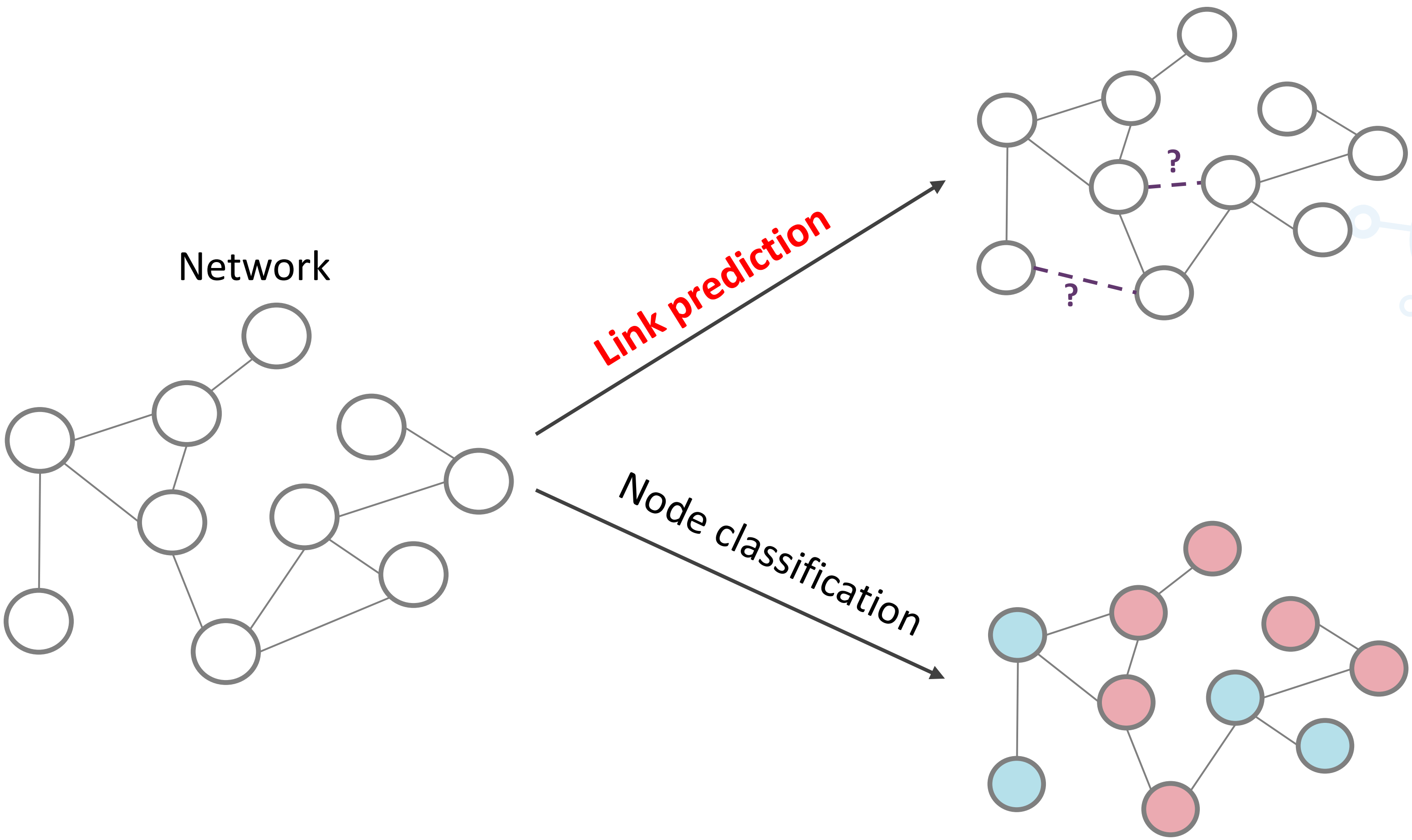
Instance: email
Feature: 'word counts'
Label: Spam or non-spam

What are some classification tasks are there in networks?

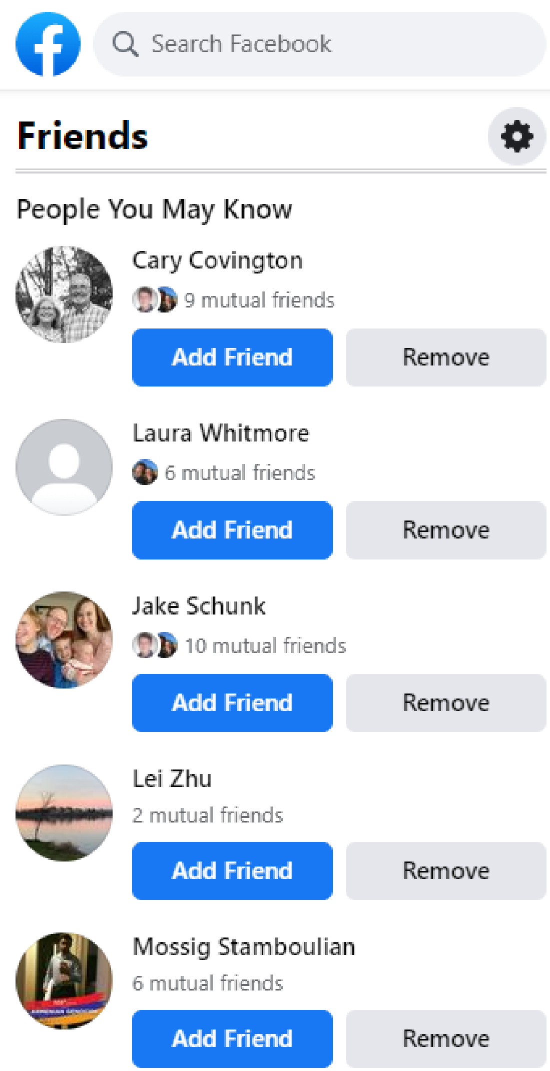
Network science

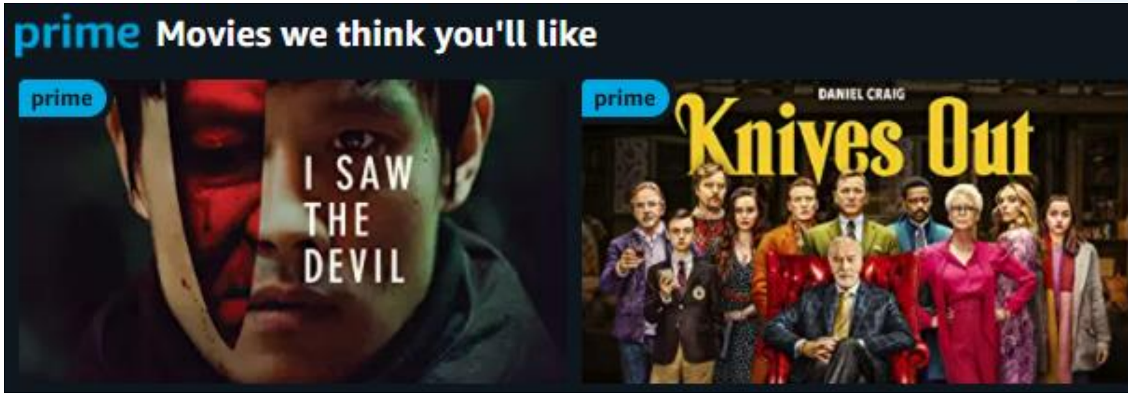


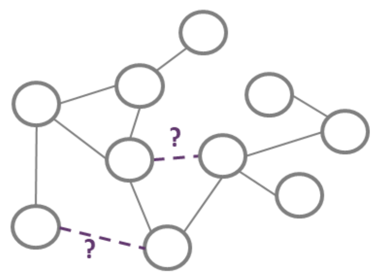
Network science



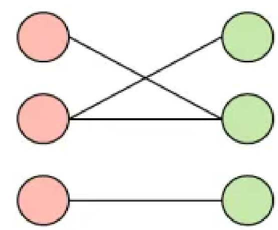
Link prediction







Facebook



Amazon

Assumption: ***similar nodes*** are likely to be connected.

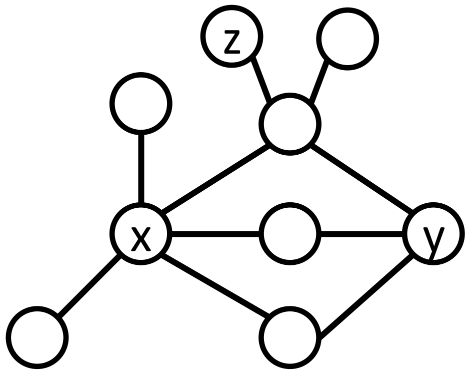
Q: How to define ***similarity***?

Similarity measures

- Common neighbors (CN)
  - Number of shared neighbors between two nodes
- Preferential attachment (PA)
  - Degree multiplication of two nodes
- ...

Dataset

	$score_{CN}$	$score_{PA}$	...
(x, y)	3	15	...
(x, z)	1	5	...
...	...	...	...



$score_{CN}(x, y) = 3$ $score_{PA}(x, y) = 15$
$score_{CN}(x, z) = 1$ $score_{PA}(x, z) = 5$

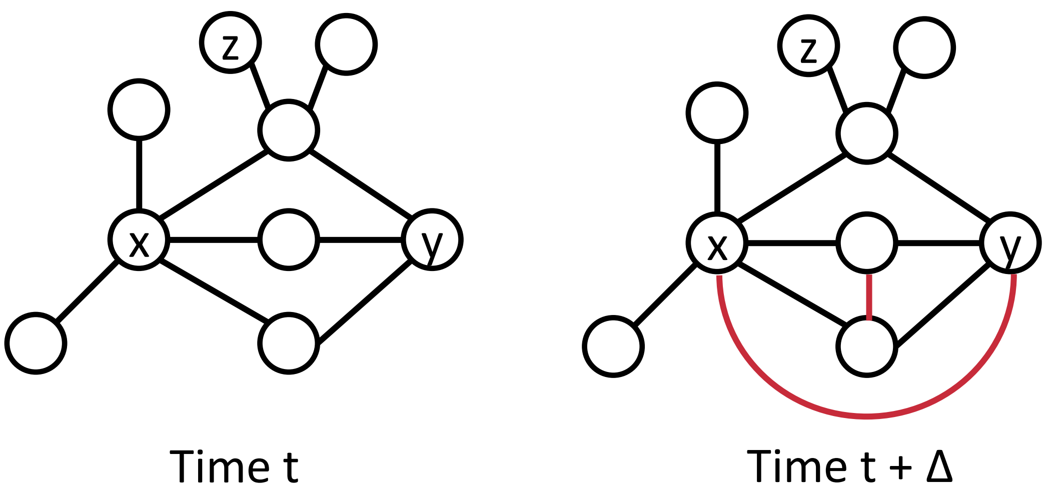
D. Liben-Nowell and J. Kleinberg, "The Link-Prediction Problem for Social Networks", JASIST 07

Supervised link prediction

Dataset

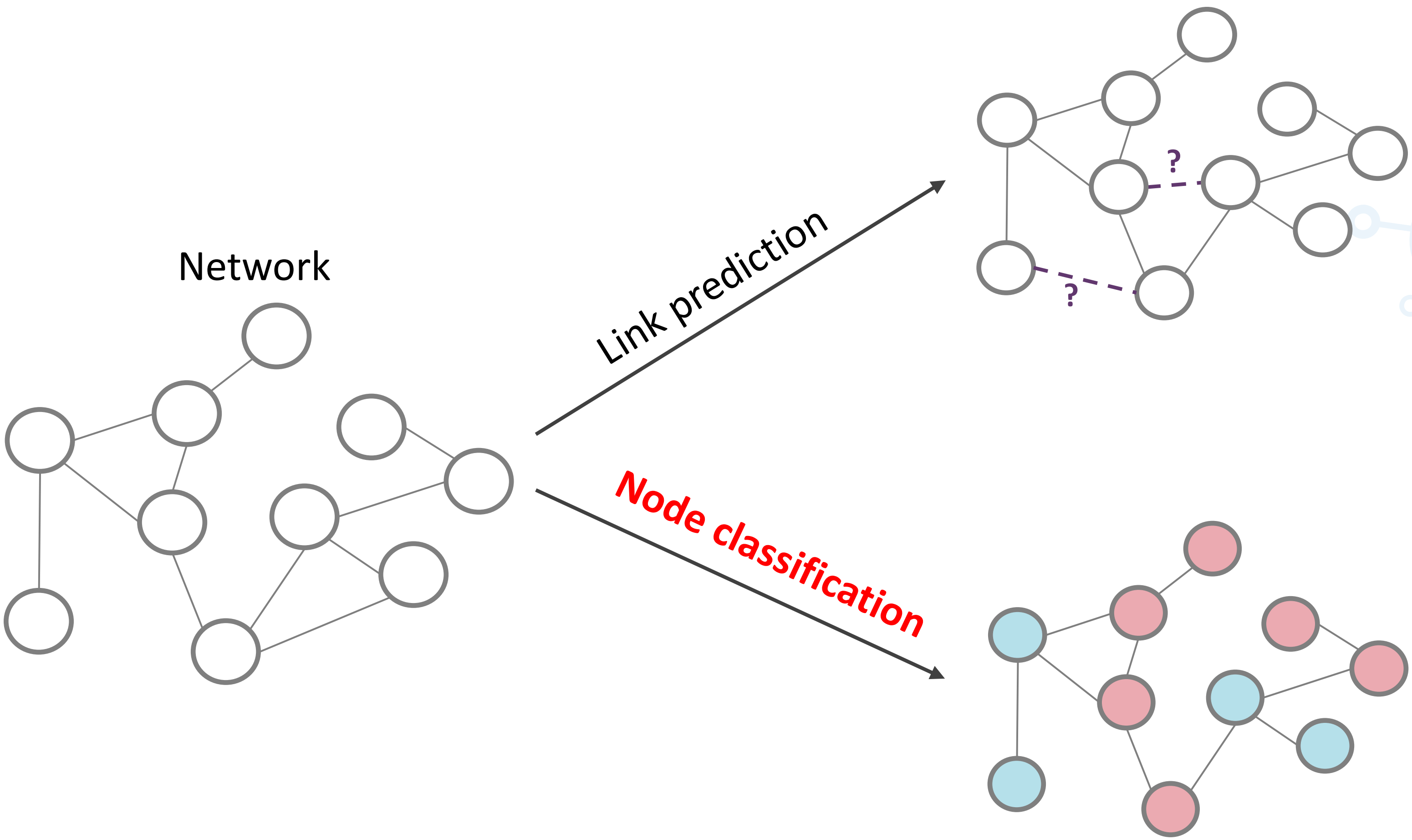
	$score_{CN}$	$score_{PA}$	…	Label
(x, y)	3	15	…	1
(x, z)	1	5	…	0
…	…	…	…	…

Binary classification!



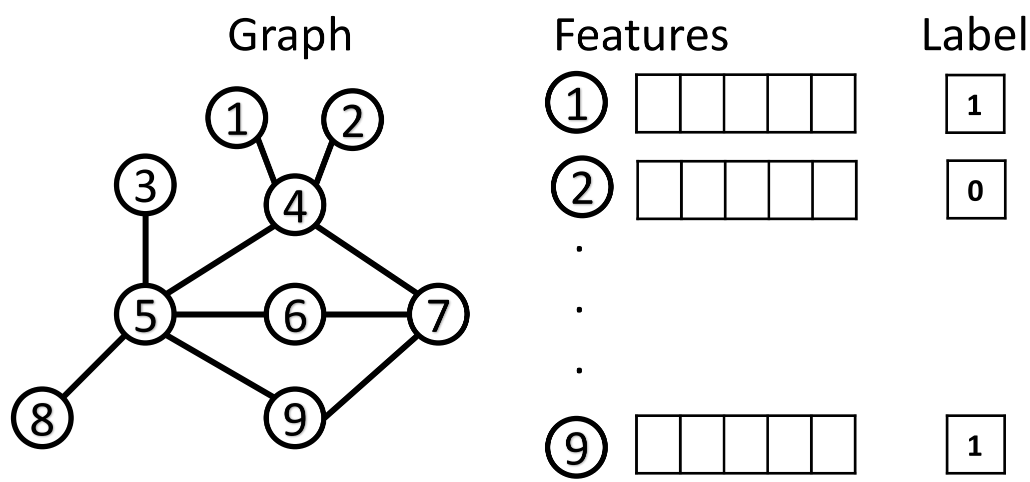
D. Liben-Nowell and J. Kleinberg, "The Link-Prediction Problem for Social Networks", JASIST 07

Network science



Node classification

- Blog catalog prediction
  - Graph: Blogs and its connection
  - Feature: Blog content
  - Node label: catalog
- COVID-19 prediction
  - Graph: Patient contact network
  - Feature: vaccinated? Immunity?
  - Node label: infection

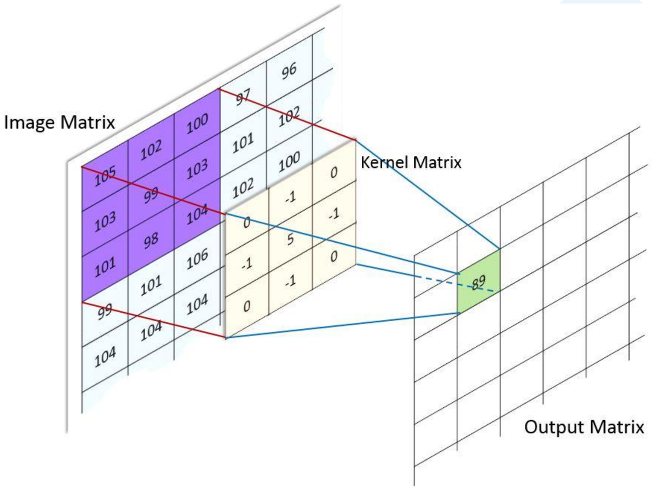


Q: How to train ML model to take into account the *connectivity*?

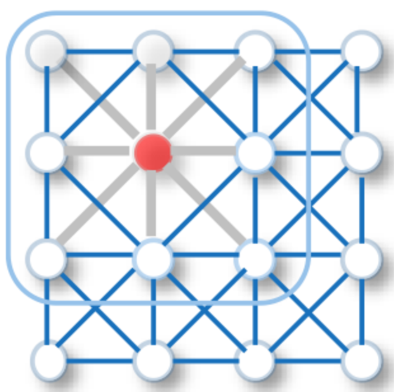
Q: Can we allow neighboring nodes features to affect each nodes' features?

Idea from CNN

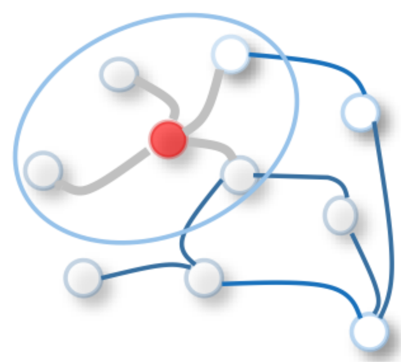
- Idea of convolutional neural network (CNN) architecture
  - Combine nearby image pixels to see a bigger picture



- Application of CNN to networks
  - Extract neighborhood information and



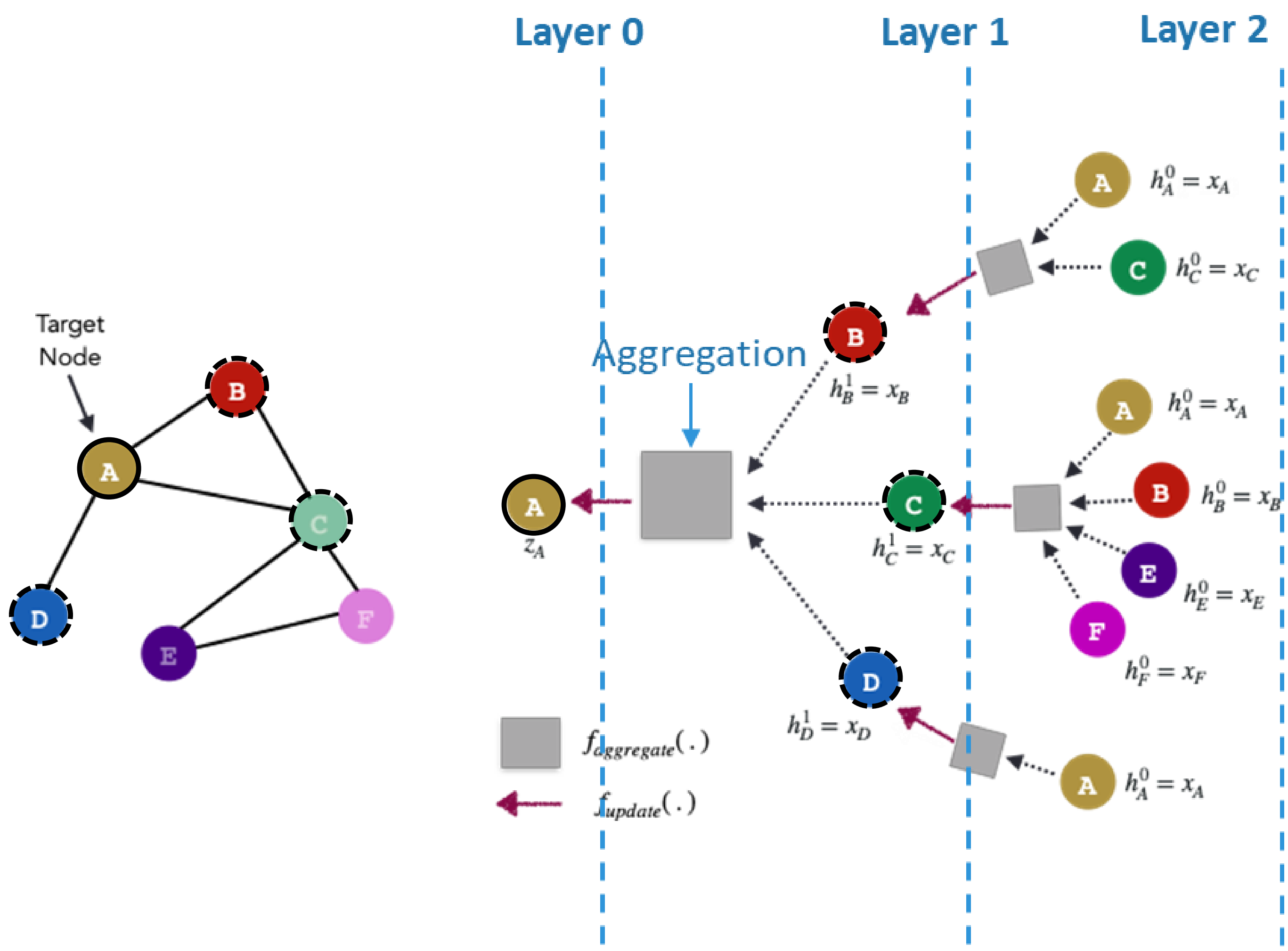
2-D convolution



Graph convolution

Bacciu, D. et al, (2020). A gentle introduction to deep learning for graphs. Neural Networks.

Graph convolutional networks



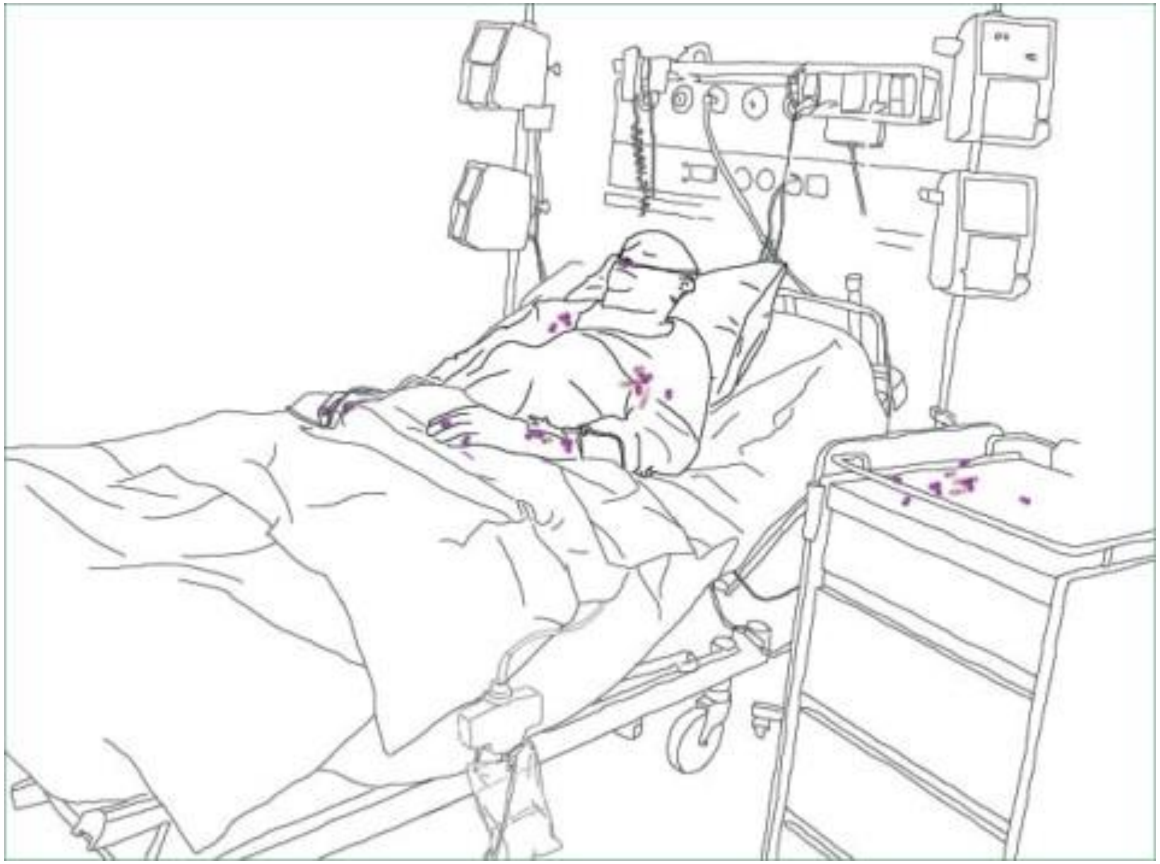
GraphSAGE model. Hamilton et al., NeurIPS 2017

Part2 Application to healthcare

Healthcare Associated Infections
- Computational Modeling and Inference

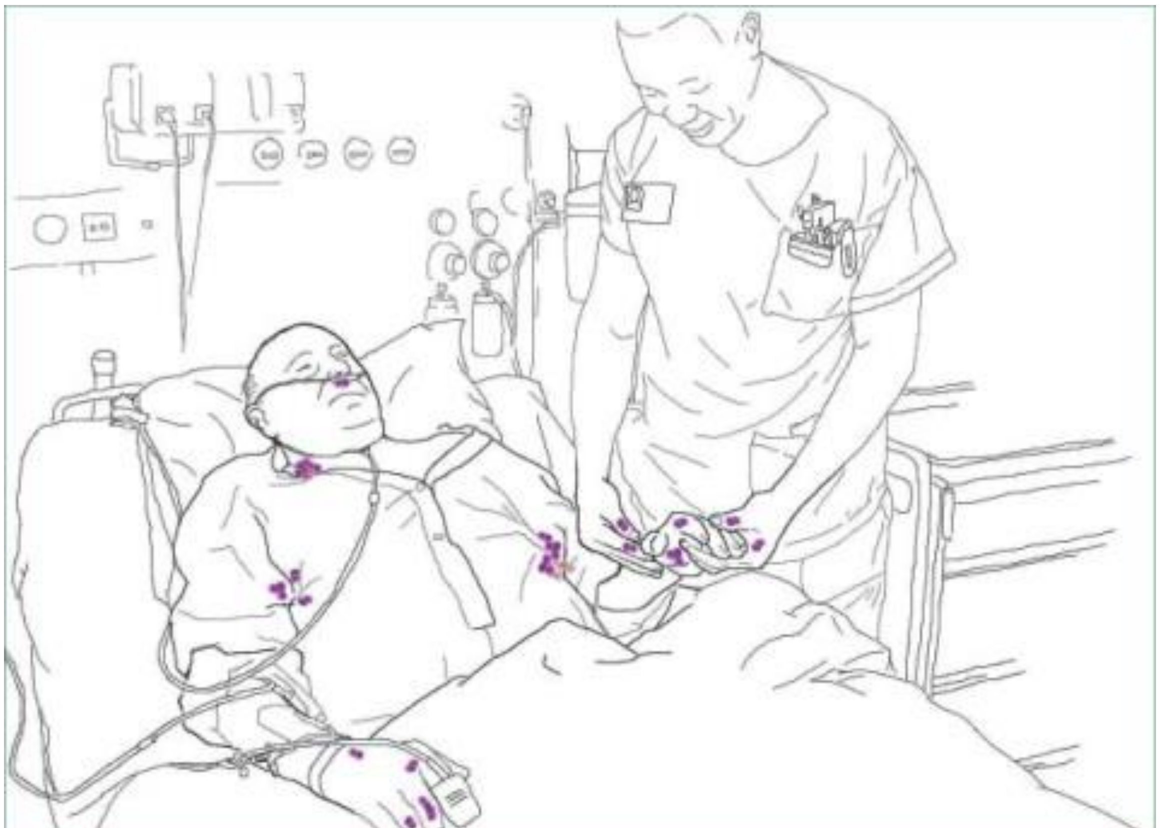
How to design interventions to reduce the spread of COVID-19 in hospital?

Healthcare associated infection (HAI)



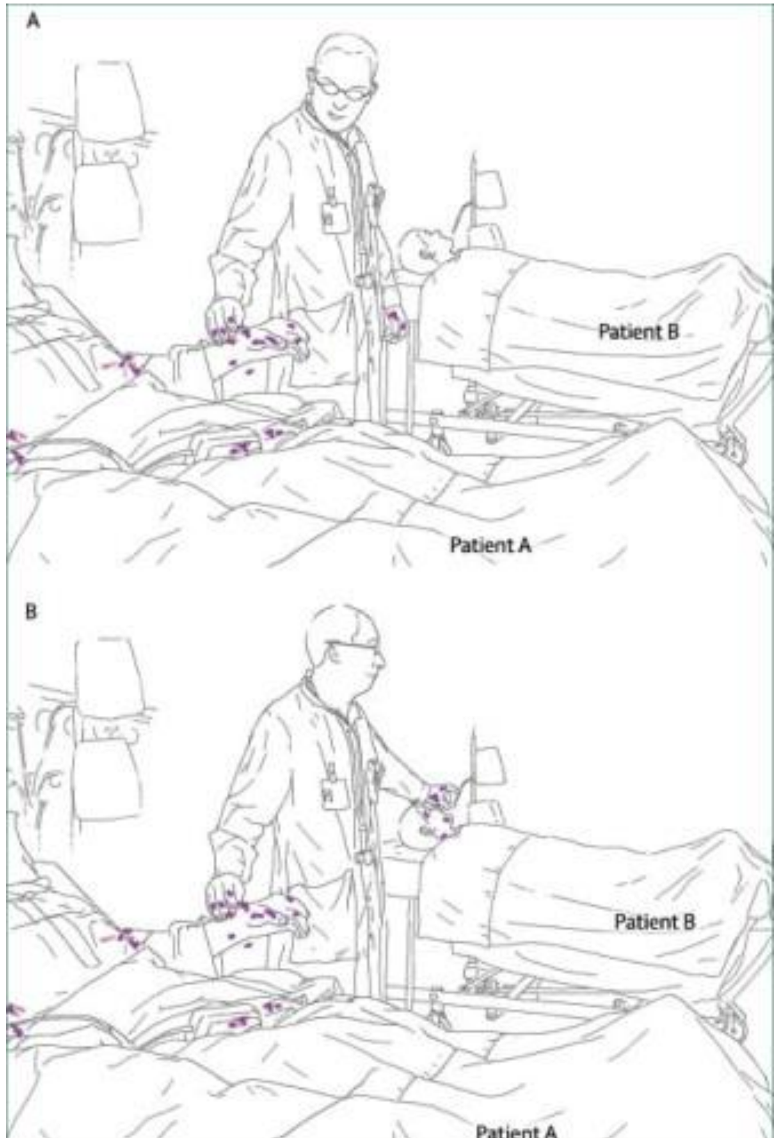
Didier Pittet et al., “Evidence-based model for hand transmission during patient care and the role of improved practices”, The Lancet Infectious Diseases, 2006

Healthcare associated infection (HAI)



Didier Pittet et al., “Evidence-based model for hand transmission during patient care and the role of improved practices”, The Lancet Infectious Diseases, 2006

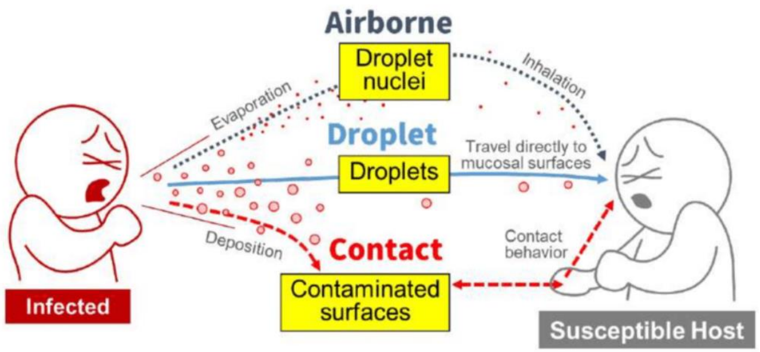
Healthcare associated infection (HAI)

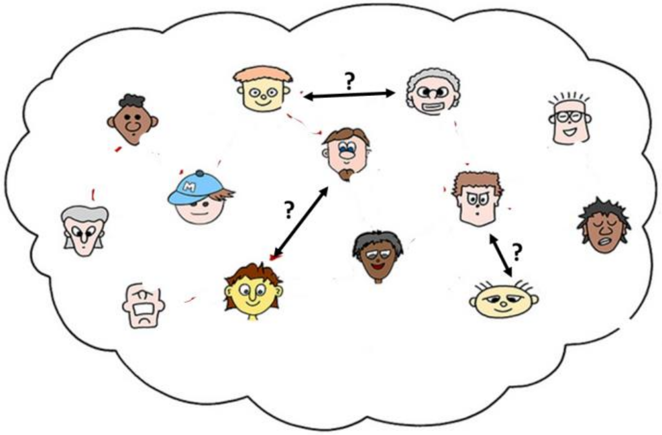


Didier Pittet et al., “Evidence-based model for hand transmission during patient care and the role of improved practices”, The Lancet Infectious Diseases, 2006

HAIs are threat to patients

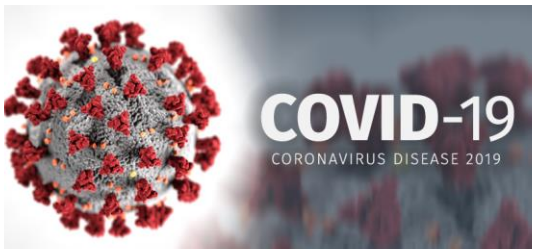
- Each year, roughly 4% of patients in the US are diagnosed with infection during their care in the hospital [*]
- Therefore, healthcare facilities are interested in preventing HAIs
- Challenges: ***Complex nature*** of disease and contacts

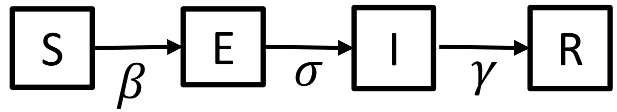


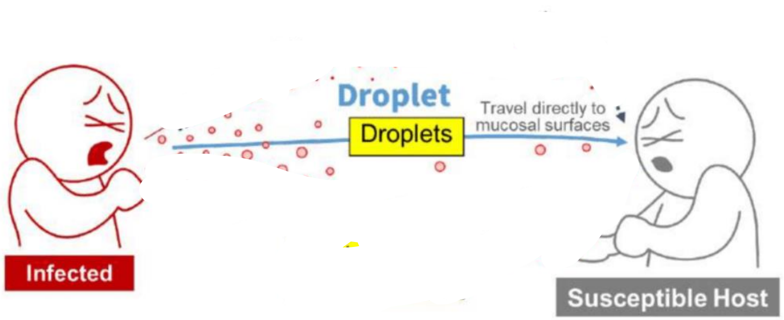


[*] Centers for Disease Control and Prevention (CDC), "Healthcare-associated infections (hais)," https://www.cdc.gov/winnablebattles/report/HAIs.html.

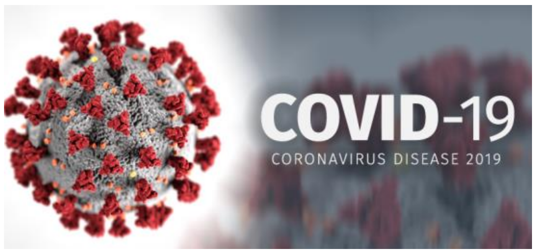
Complex disease -> compartmental model

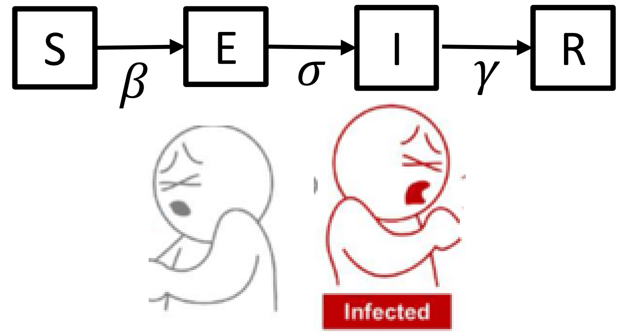


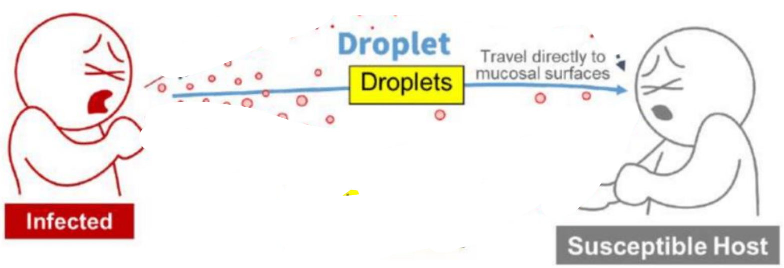




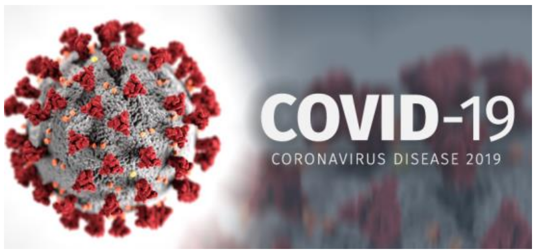
Complex disease -> compartmental model



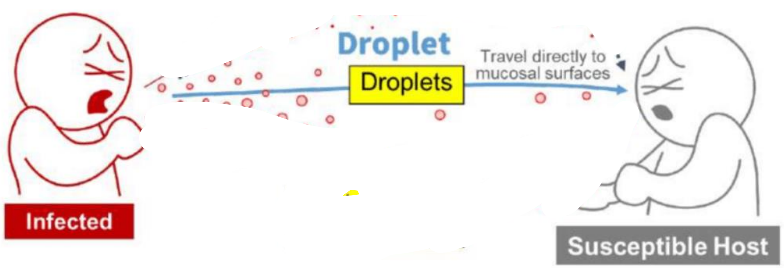




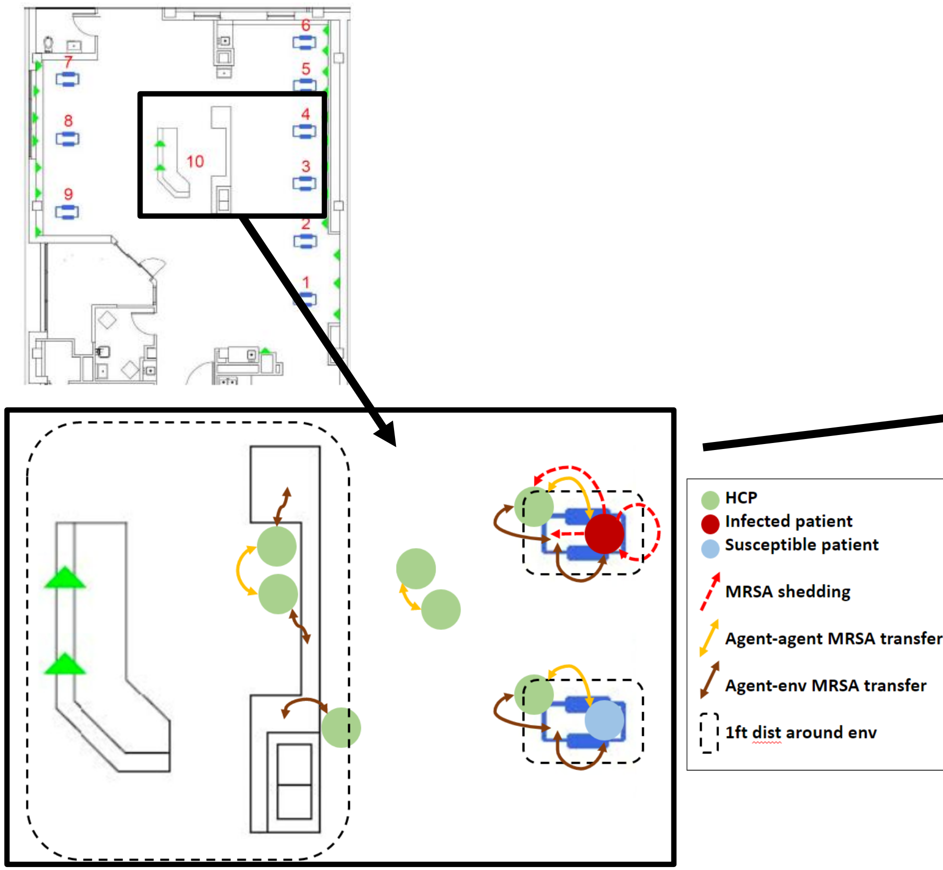
Complex disease -> compartmental model

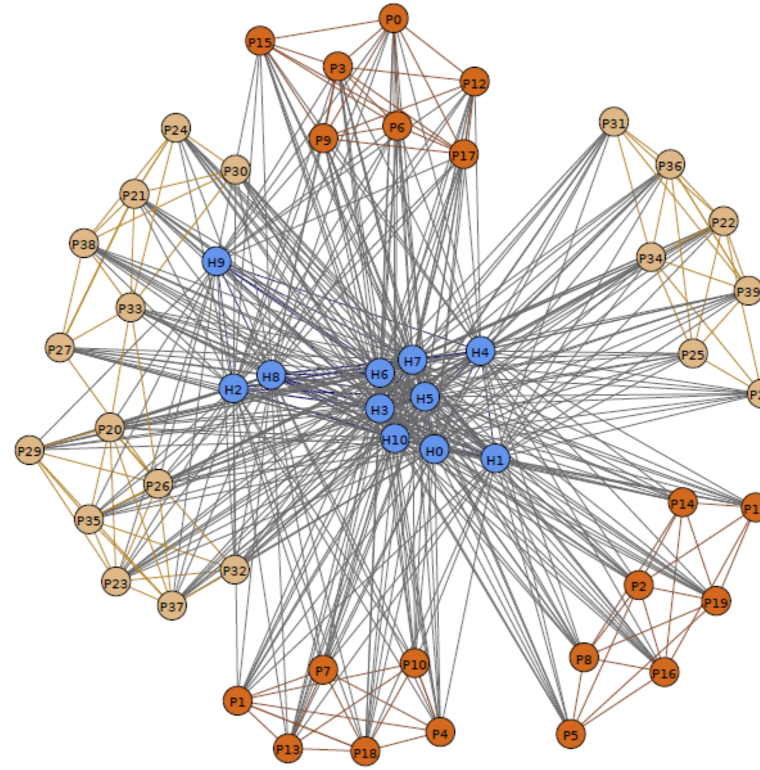


$$S \xrightarrow{\beta} E \xrightarrow{\sigma} I \xrightarrow{\gamma} R$$



Complex contacts -> contact network



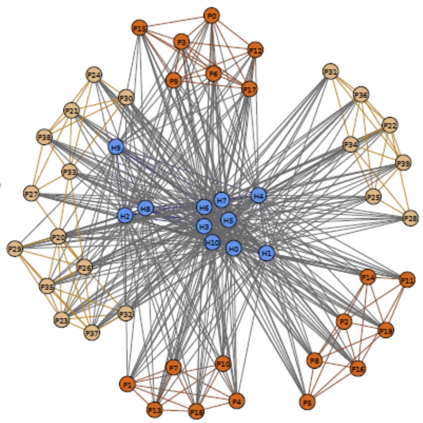


Node: Individual
Edge: Physical proximity

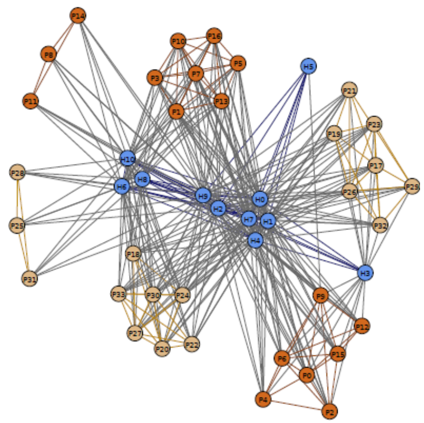
H. Jang, et al., "Evaluating Architectural Changes to Alter Pathogen Dynamics in a Dialysis Unit," **ASONAM 2019 [Best Paper Award]**

Effect of NPIs on COVID-19 shedding model

Contact network (contact if $\leq$ 6 ft)



(a) 11 HCPs, 40 patients



(b) 11 HCPs, 34 patients

Interventions

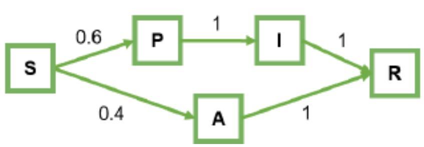
Baseline: No intervention

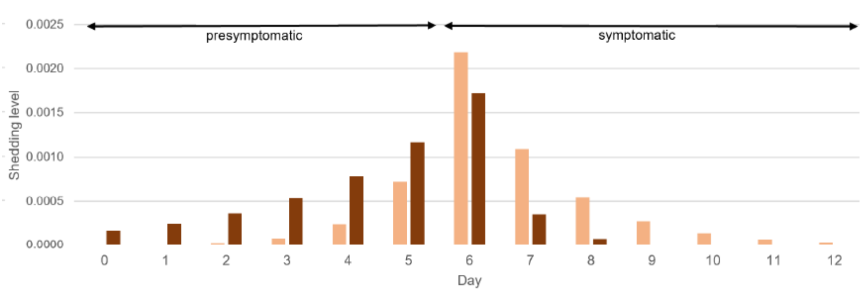
Baseline+: Surgical mask, social distancing, moving dialysis chairs apart

Baseline++: Baseline+ & infectious patient isolation, preemptive isolation of exposed HCP

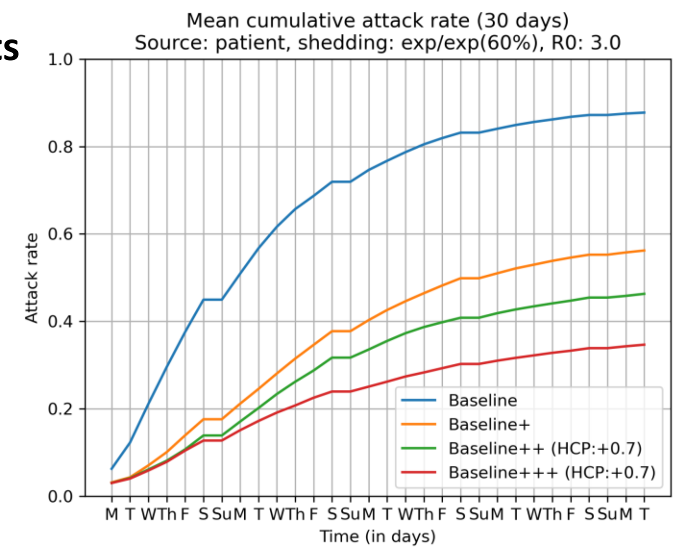
Baseline+++: Baseline++ & N95 to all HCPs for 2 weeks upon detection of the symptomatic patient

Viral shedding model





Results



[+] **H. Jang**, P. M. Polgreen, A. M. Segre, and S. V. Pemmaraju, "Covid-19 modeling and non-pharmaceutical interventions in an outpatient dialysis unit," **PLOS Computational Biology 2021**, ***under review***

How to capture medical history of patients?

Prediction tasks in healthcare

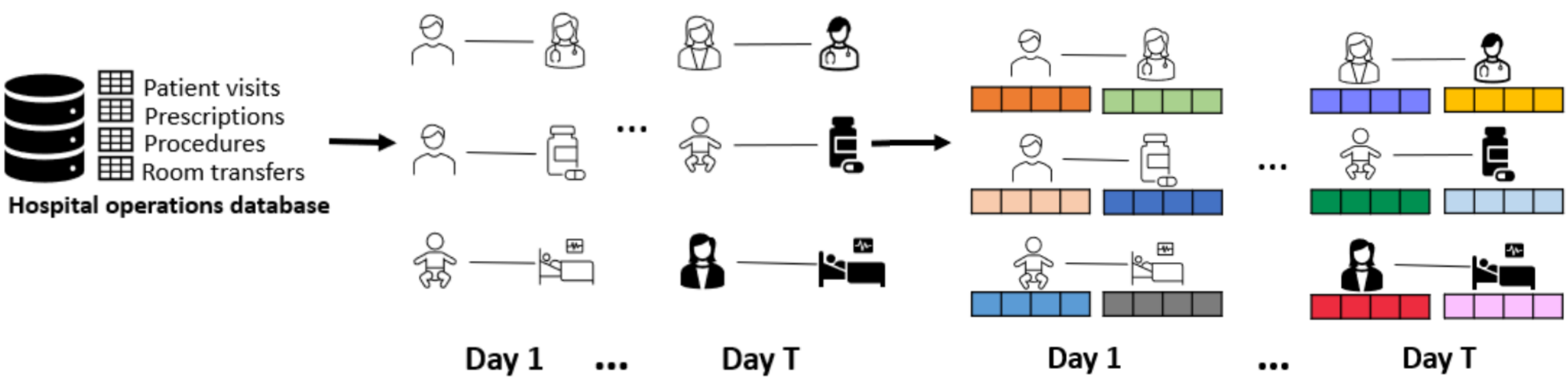
- Some patients, get infected to HAI during hospitalization
- Some has adverse events (e.g., sudden transfer into MICU)

Can we use machine learning to ***predict*** these events?

Data preparation is too costly

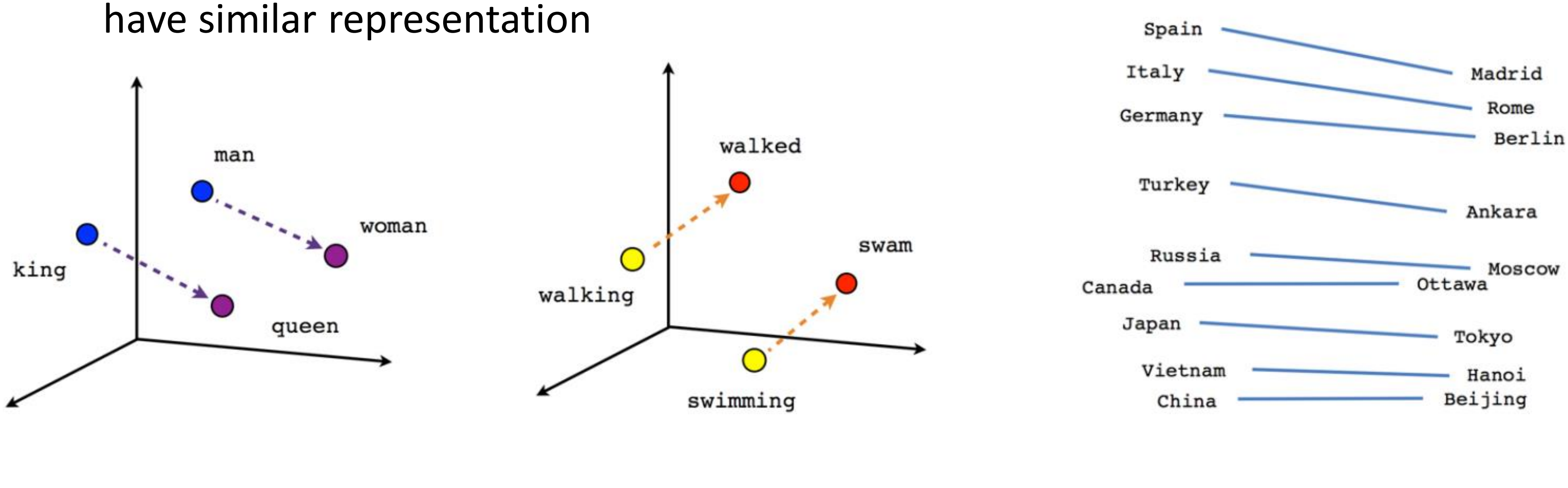
- It's costly to ***design*** and ***implement*** the data pipeline
  - Each task (e.g., HAI prediction) needs a **medical expert** for feature engineering
  - Each disease has different risk factors
  - Data scientist is needed to extract these feature from the EHR system

Can we simplify this complicated feature generation procedure?
Can we capture the medical history of the patient in an ***embedding*** for clinical decision support systems in healthcare?



Embedding in natural language processing (NLP)

- **Skipgram** [*]: ***word embedding*** is learned by maximizing the likelihood of observing co-occurring words
- Input: a document (a set of sentences)
- Task: Learn a vector representation of *word* such that nearby words would have similar representation



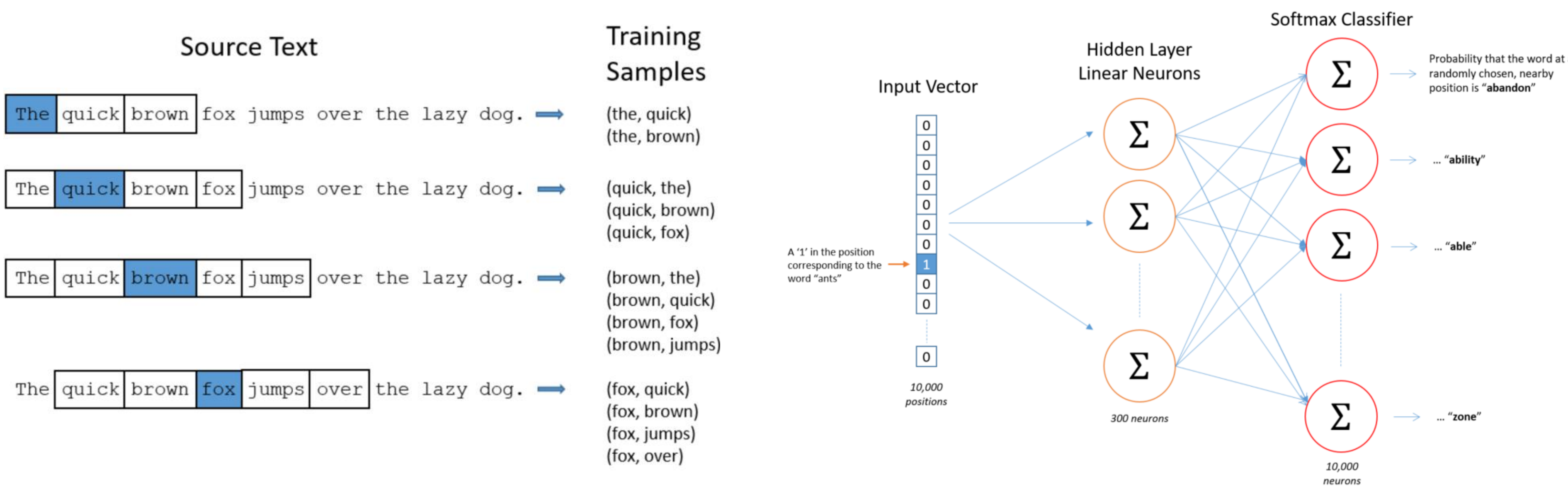
Male-Female　　Verb tense　　Country-Capital

https://towardsdatascience.com/creating-word-embeddings-coding-the-word2vec-algorithm-in-python-using-deep-learning-b337d0ba17a8

[*] T. Mikolov et al., "Distributed representations of words and phrases and their compositionality," NeurIPS 2013

Embedding in natural language processing (NLP)

- **Skipgram** [*]: ***word embedding*** is learned by maximizing the likelihood of observing co-occurring words
- Input: a document (a set of sentences)
- Task: Learn a vector representation of *word* such that nearby words would have similar representation

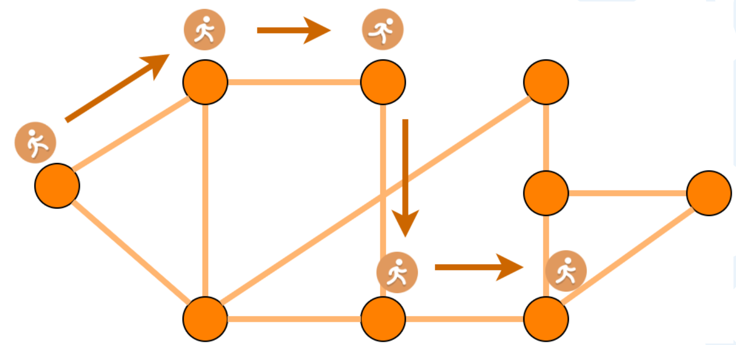


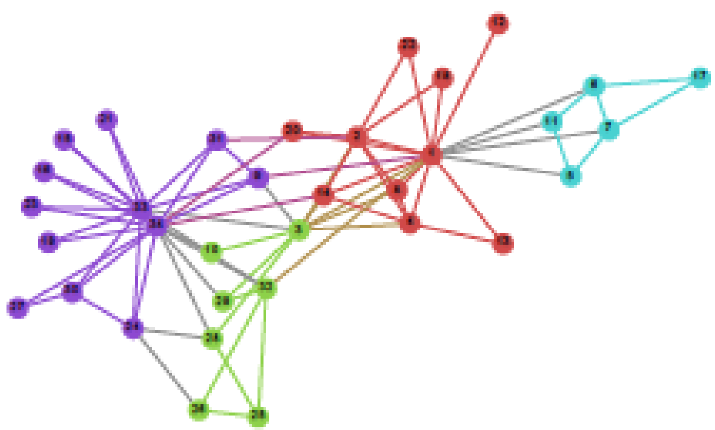
[*] T. Mikolov et al., "Distributed representations of words and phrases and their compositionality," NeurIPS 2013

Embedding in networks

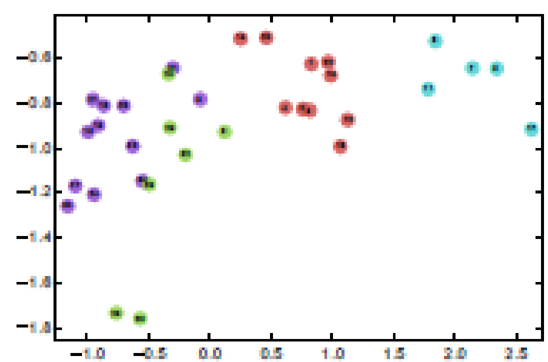
Embedding (DeepWalk)

- **DeepWalk** [-]: ***node embedding*** is learned by maximizing the likelihood of observing nearby nodes
  - Graph (= document)
  - Short random walks (= sentence)
  - Node (= word)





(a) Input: Karate Graph



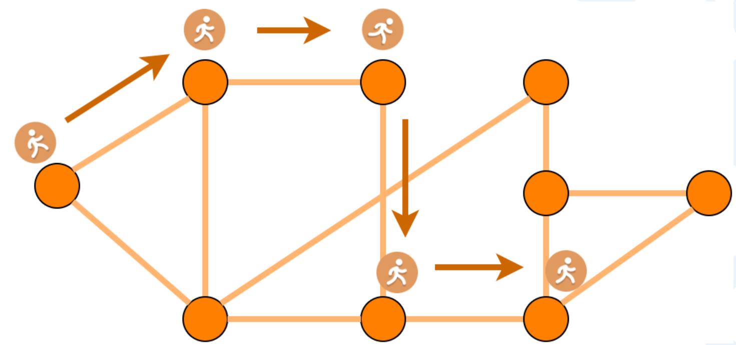
(b) Output: Representation

[-] B. Perozzi, R. Al-Rfou, S. Skiena, "Deepwalk: Online learning of social representations", KDD 14

Embedding in networks

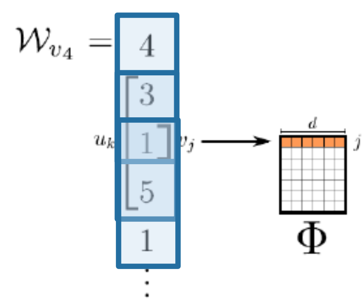
Embedding (DeepWalk)

- **DeepWalk** [-]: ***node embedding*** is learned by maximizing the likelihood of observing nearby nodes
  - Graph (= document)
  - Short random walks (= sentence)
  - Node (= word)



$$\mathcal{W}_{v_4} = v_4 \rightarrow v_3 \rightarrow {\color{red}v_1} \rightarrow v_5 \rightarrow v_1 \rightarrow v_{46} \rightarrow v_{51} \rightarrow v_{89}$$

- Step1: Generate short ***random walks*** for each node in the graph
- Step2: Prepare pair of nearby nodes
- Step3: Train **Skip-gram**



- Map the vertex under focus (${\color{red}v_1}$) to its representation.
- Define a window of size w
- If w = 1 and v = ${\color{red}v_1}$

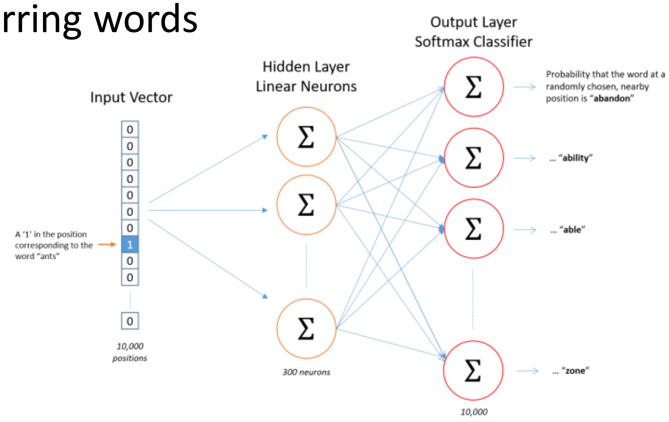
Maximize: $\Pr(v_3|\Phi({\color{red}v_1}))$

$\Pr(v_5|\Phi({\color{red}v_1}))$

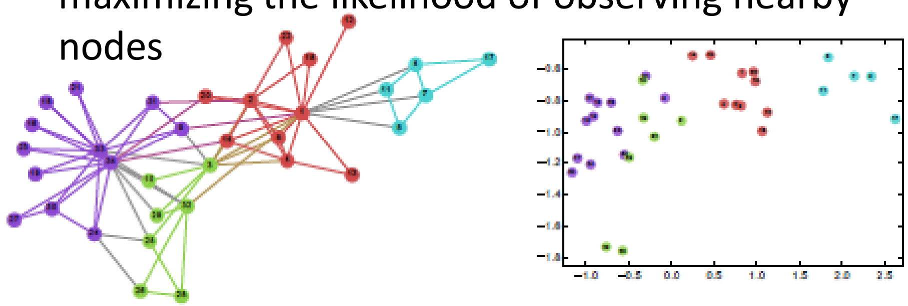
[-] B. Perozzi, R. Al-Rfou, S. Skiena, "Deepwalk: Online learning of social representations", KDD 14

patient embedding

Skipgram [*]: ***word embedding*** is learned by maximizing the likelihood of observing co-occurring words

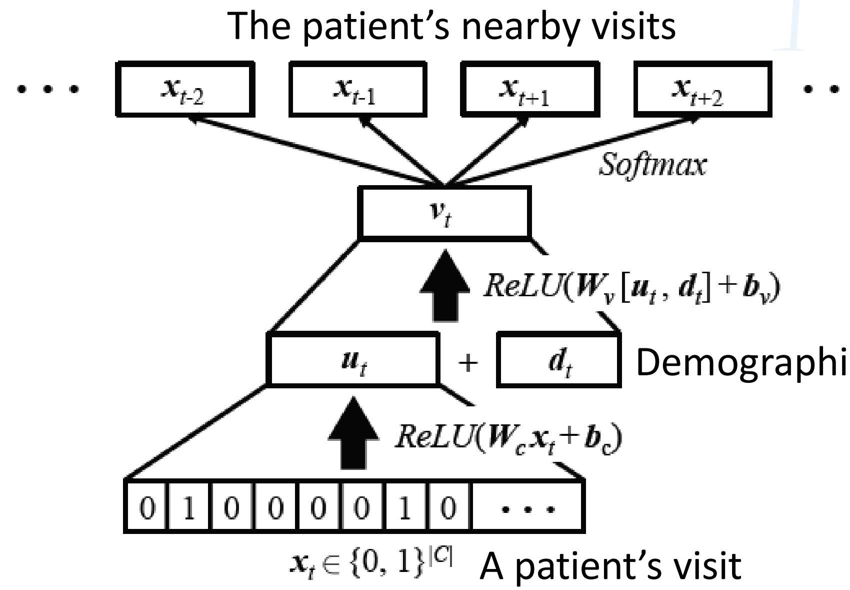


DeepWalk [-]: ***node embedding*** is learned by maximizing the likelihood of observing nearby nodes



(a) Input: Karate Graph (b) Output: Representation

Med2Vec [+]: ***patient visit embedding*** is learned by maximizing the likelihood of observing nearby visits

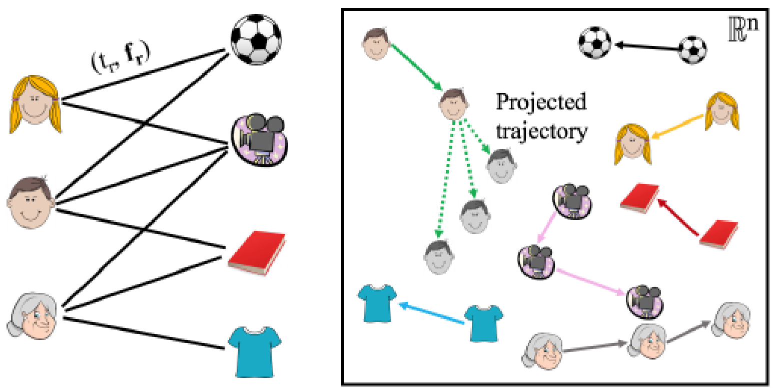


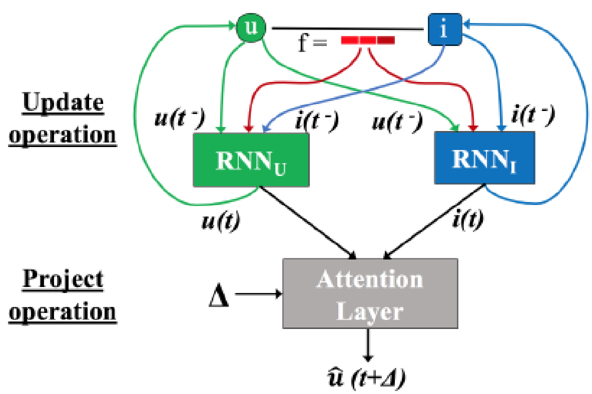
Skipgram	DeepWalk	Med2Vec
Word	Node	Patient visit
Sentence	Random walk	Nearby visit

[*] T. Mikolov et al., "Distributed representations of words and phrases and their compositionality," NIPS 2013
[-] B. Perozzi et al., "DeepWalk: online learning of social representations," KDD '14
[+] E. Choi et al., "Multi-layer representation learning for medical concepts," KDD '16

patient embedding (dynamic)

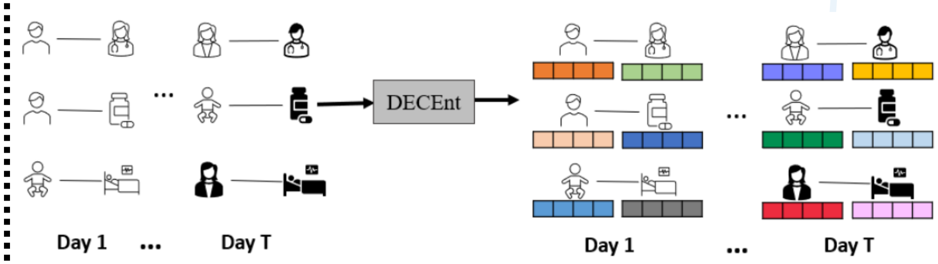
JODIE [*]: Learns user embedding and item embedding over time based on interactions





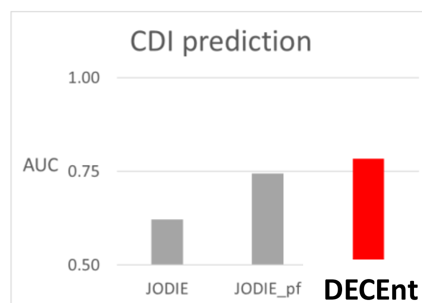
Next item prediction

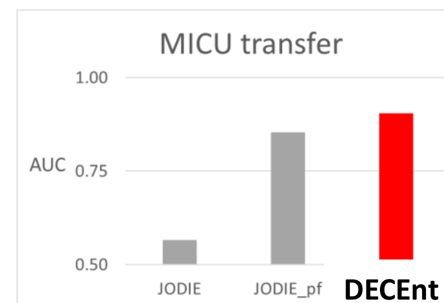
DECEnt [+]: Learns patient embedding and {doctor, medication, room} embedding over time



Next interaction prediction

For each interaction (e.g., patient-doctor) **DECEnt** predicts **next interaction** (e.g., doctor encounter)





[*] S. Kumar, X. Zhang, and J. Leskovec, "Predicting dynamic embedding trajectory in temporal interaction networks,", KDD 19

[+] **H. Jang**, S. Lee, H. Hasan, P. Polgreen, S. Pemmaraju, B. Adhikari, "Dynamic Healthcare Embeddings for Improving Patient Care", ***in submission to*** **CIKM '21**

Q / A

Email: jhkmath@gmail.com

Thank you!

Email: jhkmath@gmail.com