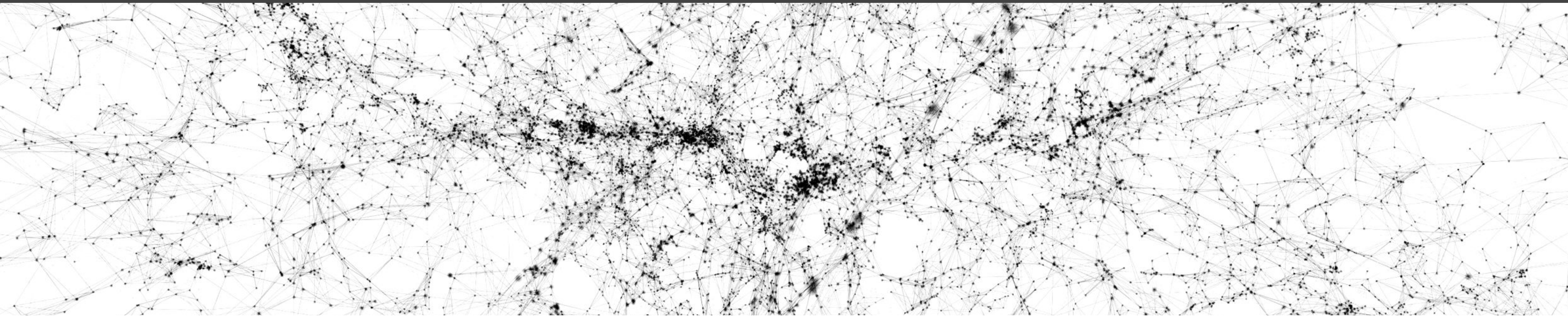
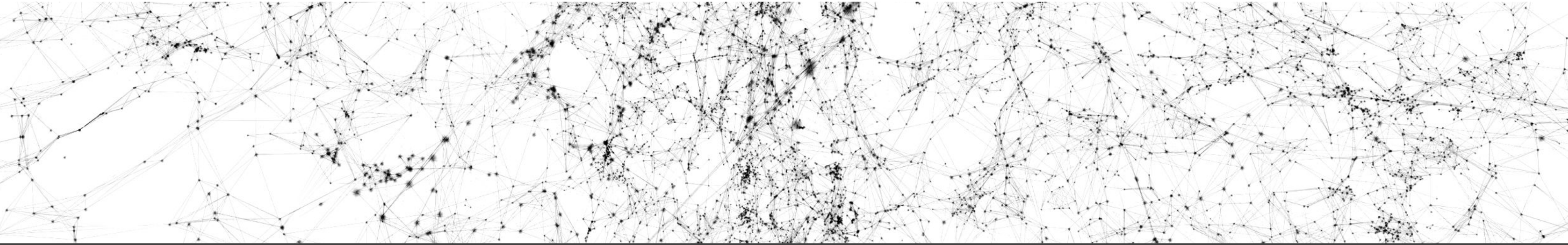


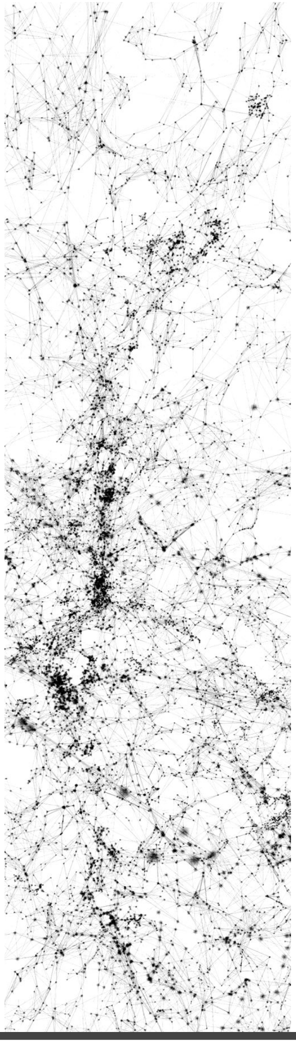
Background

Analysis of complex interconnected data



Outline

- **Quick Recap of important points**
  - General info & Grading
  - Who is in the class
- Learning the vocabulary of Network Science
  - Evolution of the field and scale of the data
  - Types of Networks: simple, directed, temporal, bipartite, etc
  - Adjacency matrix, powers of A, Laplacian matrix
  - Shortest path, connectivity, connected components, GCC



Grading details

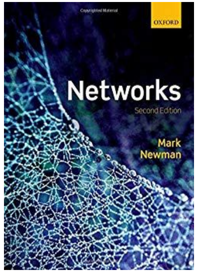
- 40% project (10% proposal, 10% progress report, 20% final report)
- 30% assignments (3x10%)
- 25% presentations of assigned papers
- 5% reviewing assignments
  note: most of the grading is by peer-assessment
- bonus points:
  - 5 point for the best class presentation
  - 5 points for the best project proposal
  - 5 point for the best reviewer
  - 10 points for the best project
  - 1 point for each interesting point you share at the end of a class from the readings (for the current or previous lectures) which was not covered in the class

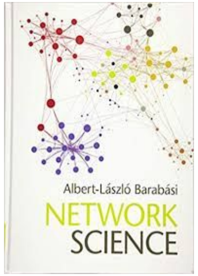
Schedule

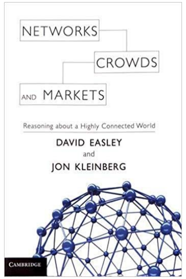
Wed., Sep. 2
~~Mon., Sep. 7~~
Wed., Sep. 9
  Topic: Background, slides
  Reading: NI chapter 6, NS chapter 2
Mon., Sep. 14
Wed., Sep. 16
Mon., Sep. 21

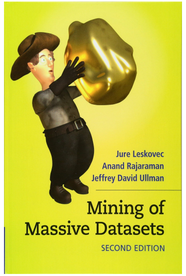
Reference Materials

- **Main textbooks**
  - **[NI] Networks: An Introduction** by M.E.J. Newman, ebook at library
  - **[NS] Network Science** by Albert-Barabasi, available online
- **Other textbooks**
  - **Networks, Crowds and Markets** by D. Easley and J. Kleinberg, available online
  - **Mining of Massive Datasets** by Jure Leskovec, Anand Rajaraman, Jeff Ullman, available online
- **Surveys and conference papers**
  - Web (WebConference, WSDM, ICWSM), Data (KDD, ICDM, SDM, ECML/PKDD, PAKDD), Learning (ICML, NeurIPS), Networks (ASONAM, NetSci, Complex Networks), …









Announcement: first assignment out

http://www.reirab.com/Teaching/NS20/Assignment_1.pdf

Check the description, and partner up

Deadlines

- assignment 1 due on Sep. 20th
- assignment 2 due on Oct. 4th
- assignment 3 due on Oct. 18th
- project proposal slides due on Oct. 25th
- project proposal due on Nov. 1th
- Reviews (first round) due on Nov. 8th
- project progress report due on Nov. 22nd
- Reviews (second round) due on Nov. 29th
- project final report slides due on Dec. 1st
- project final report due on Dec. 6th
- Reviews (third round) due on Dec. 13th
- project revised report and rebuttal due on Dec. 20th
- note: dates are tentative, please check them for the updated deadlines

Projects - updated plan

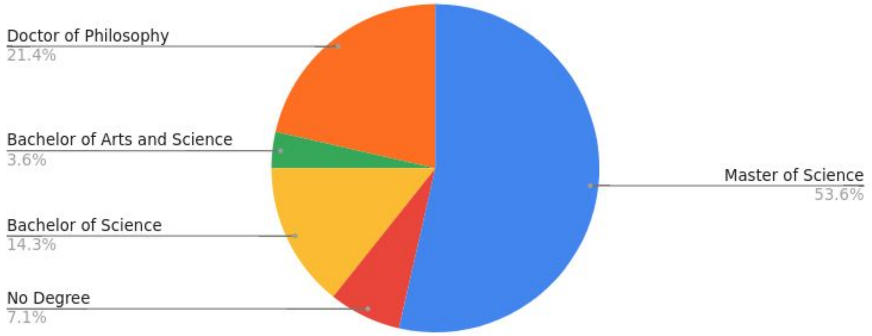
Proposals are individually

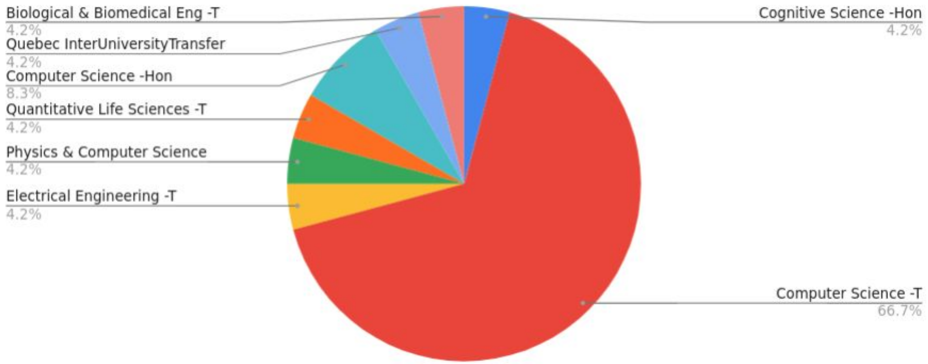
After proposal presentations, you can decide to join another project and continue as a group of two or complete the project individually

Class composition

Quick round of introductions

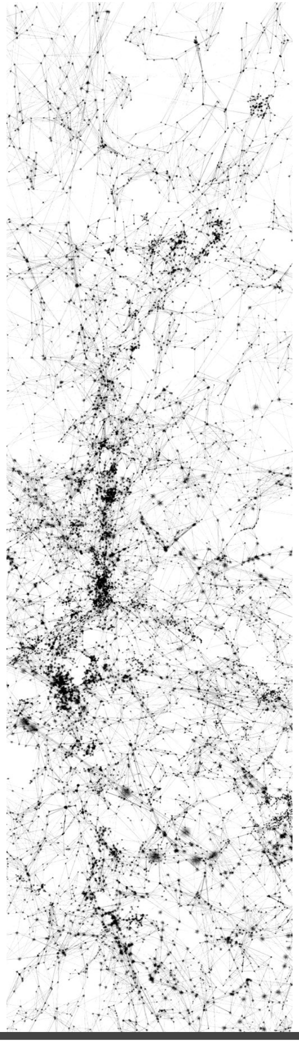
- Name
- Your background
- Any particular reason for taking this class





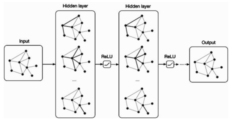
Outline

- Quick Recap of important points
  - General info & Grading
  - Who is in the class
- **Learning the vocabulary of Network Science**
  - Evolution of the field and scale of the data
  - Types of Networks: simple, directed, temporal, bipartite, etc
  - Adjacency matrix, powers of A, Laplacian matrix
  - Shortest path, connectivity, connected components, GCC



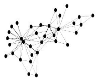
Timeline of notable works in network science

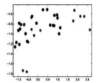
Recent Trend: **Deep Learning for Graphs**



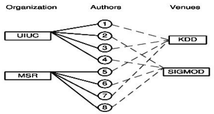
2015~2019
- **Graph Neural Networks**
- Deep Learning for Networks
- High-Order Networks [Benson et al.]

21st Century: **More CS**





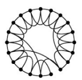
2010~2014
- Info. vs. Social Networks (Twitter) [Kwak et al.]
- **Signed** Networks [Leskovec et al.]
- Semantic Social Networks [Tang et al.]
- Four Deg. Of Separation [Backstrom et al.]
- Structural Diversity [Ugander et al.]
- Computational Social Science [Watts]
- **Network Embedding** [Perozzi et al.]

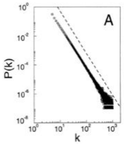


2005~2009
- Graph Evolution [Leskovec et al.]
- 3 Deg. Of Influence [Christakis & Fowler]
- Social **Influence** Analysis [Tang et al.]
- Six Deg. Of Separation [Leskovec & Horvitz]
- Network **Heterogeneity** [Sun & Han]
- Network **Embedding** [Tang & Liu]
- Computer Social Science [Lazer et al.]

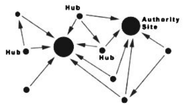
2000~2004
- Influence Max'n [Domingos & Kempe et al.]
- **Community Detection** [Girvan & Newman]
- Network Motifs [Milo et al.]
- Link Prediction [Liben-Nowell & Kleinberg]

Late 20th Century: **CS & Physics**

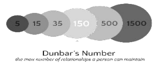




1998~1999
- **Small Worlds** [Watts & Strogatz]
- **Scale Free** [Barabasi & Albert]
- **Power Law** [Faloutsos x3]

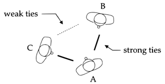


1997
- **HITS** [Kleinberg]
- **PageRank** [Page & Brin]
- Hyperlink Vector Voting [Li]

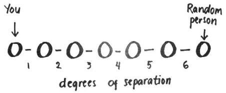


1992
- Structural Hole [Burt]
- **Dunbar's Number** [Dunbar]

20th Century: **Sociology**

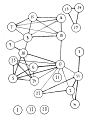


1970s
- The Strength Of **Weak Tie** [Granovetter]



1960s
- **Small Worlds** [Migram]

1950s
- **Random Graph** [Erdos, Renyi, Gilbert]
- Degree Sequence [Tuttle, Havel, Hakami]
- **Homophily** [Lazarsfeld & Merton]
- Balance Theory [Heider et al.]



1930s
- **Sociogram** [Moreno]

Graph theory is older than network science

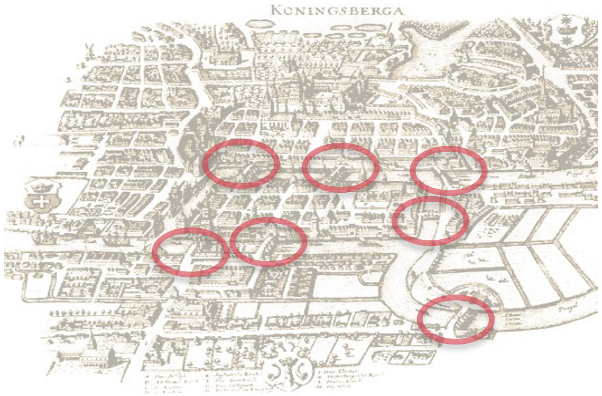
Based on Slides from Jie Tang

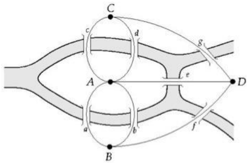
Graph Theory & Network Science

Network science borrows many concepts/theories from graph theory. The focus, however, is on **real world** graphs which have specific characteristics, and are different than random graph families commonly studied in math.

for example, regular graphs (same degree for all nodes), are irrelevant here.

Can one walk across the seven bridges and never cross the same bridge twice?





1735: **Euler's theorem:**

If a graph has more than two nodes of odd degree, there is no path. If a graph is connected and has no odd degree nodes, it has at least one path.

From Barbasi's slides

Real world graphs are Large Scale

facebook

- 2 billion MAU
- 26.4 billion minutes/day

Alibaba Group 阿里巴巴集团

- >777 million trans. (alipay)
- 200 billion on 11/11

twitter

- 320 million MAU
- Peak: 143K tweets/s

WeChat

- QQ: 860 million MAU
- WeChat: 1.1 billion MAU

Instagram

- 700 million MAU
- 95 million pics/day

Snapchat

- 300 million MAU
- 30 minutes/user/day

Based on Slides from Jie Tang

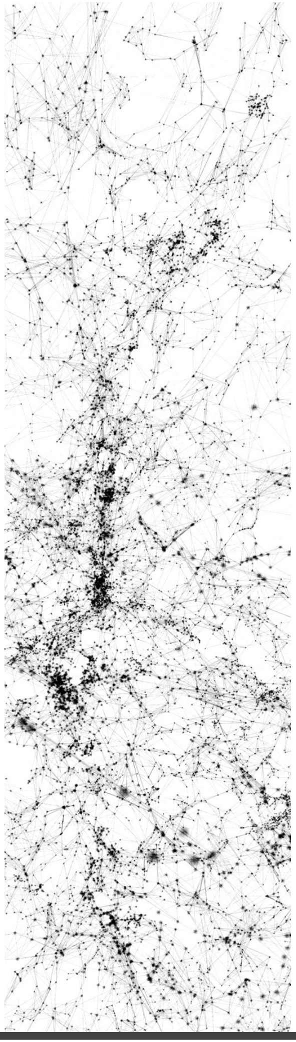
Example benchmark datasets

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L
Internet	Routers	Internet connections	Undirected	192,244	609,066
WWW	Webpages	Links	Directed	325,729	1,497,134
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826
Email	Email addresses	Emails	Directed	57,194	103,731
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908
Citation Network	Paper	Citations	Directed	449,673	4,689,479
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930

You can download these bundled from Barbasi's website, for the assignment

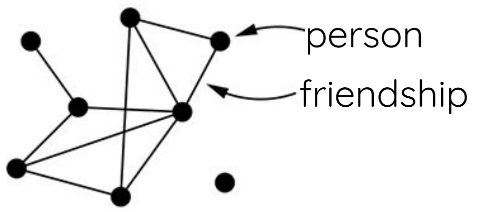
Outline

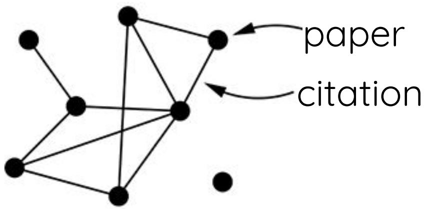
- Quick Recap of important points
    - General info & Grading
    - Who is in the class
- **Learning the vocabulary of Network Science**
    - Evolution of the field and scale of the data
    - **Types of Networks: simple, directed, temporal, bipartite, etc**
    - Adjacency matrix, powers of A, Laplacian matrix
    - Shortest path, connectivity, connected components, GCC



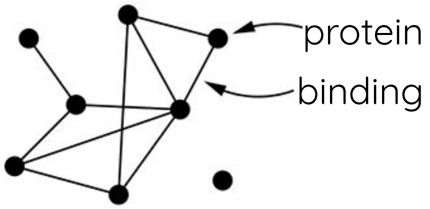
Interconnected Data as Graphs

- Nodes (or Vertices)
  - Proteins, Neurons, People
- Edges (or Links)
  - interactions, friendships





- Two vertices are adjacent if they share a common edge
- Two adjacent vertices are neighbors
- An edge is incident with another edge if they share a vertex
- An edge is incident with two vertices



Adjacency: the default data structure

Adjacency Matrix

$$
\begin{array}{c|cccccccccccc|}
 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
2 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
3 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
4 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
5 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
6 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
7 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
8 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\
9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
10 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\
11 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0
\end{array}
$$

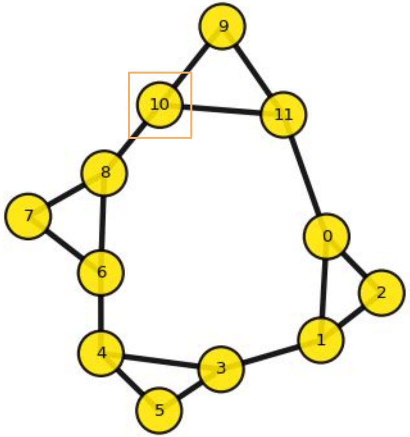
Adjacency List

0 : { 1 , 2 , 11 }
1 : { 0 , 2 , 3 }
2 : { 0 , 1 }
3 : { 1 , 4 , 5 }
4 : { 3 , 5 , 6 }
5 : { 3 , 4 }
6 : { 4 , 7 , 8 }
7 : { 6 , 8 }
8 : { 6 , 7, 10 }
9 : { 10 , 11 }
10 : { 8 , 9 , 11 }
11 : { 0 , 9 , 10 }

Edge List

{ (0, 1), (0, 2), (0, 11),
(1, 0), (1, 2), (1, 3),
(2, 0), (2, 1),
(3, 1), (3, 4), (3, 5),
(4, 3), (4, 5), (4, 6),
(5, 3), (5, 4),
(6, 4), (6, 7), (6, 8),
(7, 8), (7, 6),
(8, 6), (8, 7), (8, 10)
(9, 10), (9,11),
(10, 8), (10, 9), (10, 11),
(11, 0), (11, 9), (11, 10) }

Simple Graph

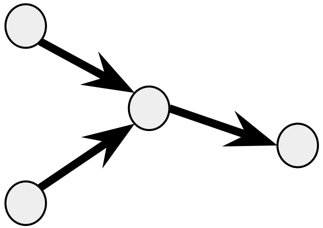


$$G(V,E), V = \{1 \ldots n\}, E = \{(i,j) | i,j \in [1 \ldots n]) \wedge A_{ij} = 1\}$$

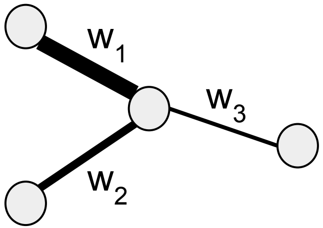
Real world graphs are sparse (lots of zeros) and we use sparse matrix representations which in practice are similar to adjacency (LIL format)/edge list (COO format) and only store non-zero values.

Not Simple

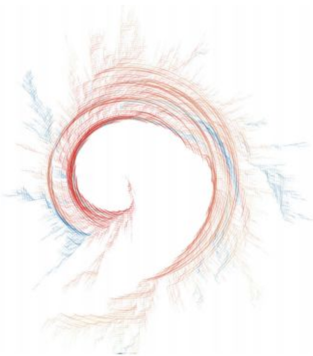
- Directions
  - E.g. who follows who at Twitter

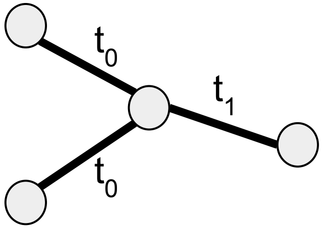


- Weights
  - E.g. friendship strength, or travel cost



- Time
  - E.g. your friendships changes



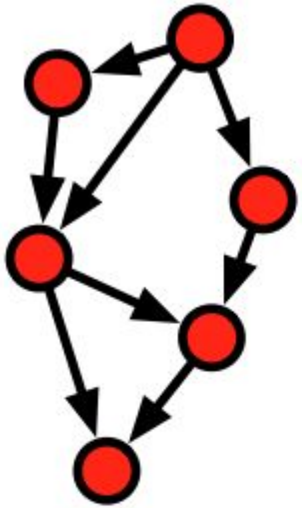


Directed Networks Examples

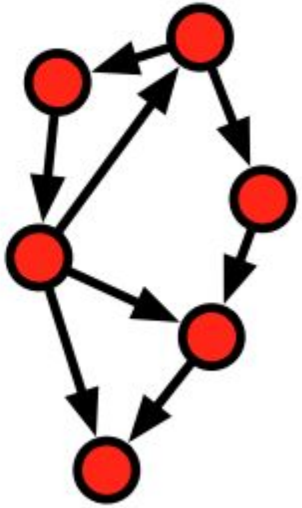
citation networks

foodwebs*

epidemiological



directed acyclic graph



directed graph

WWW

friendship?

flows of goods, information

economic exchange

dominance

neuronal

transcription

time travelers

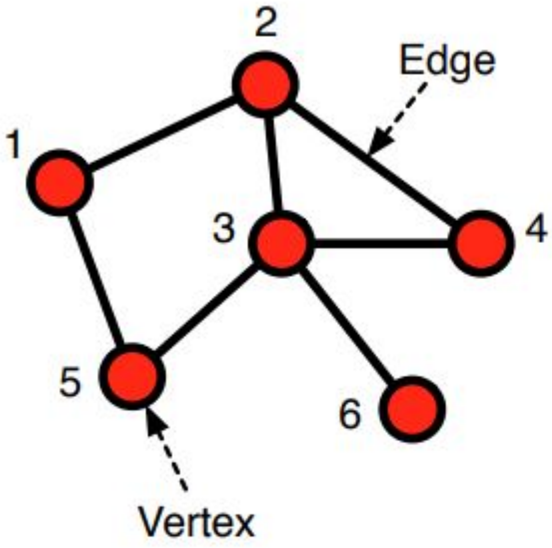
[From Clauset's slides]

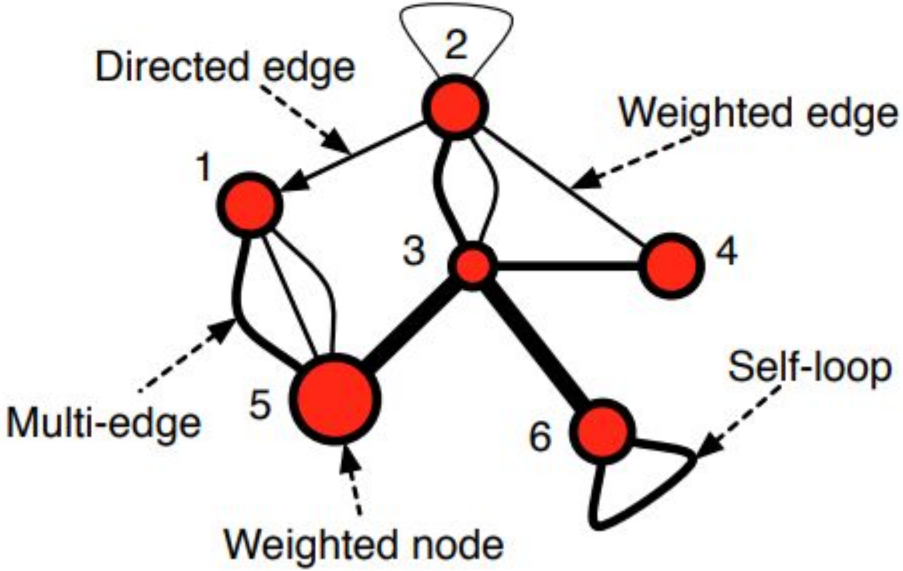
Adjacency Matrix

- Symmetric if graph is undirected
  - **$A_{ij} = A_{ji}$**
- Directected, not symmetric
  - **$A_{ij} \neq A_{ji}$**
- Weighted, not binary
  - **$[0,1] \Rightarrow R^+$**
- Temporal
  - **Matrix ⇒ Tensor**

$$
\begin{array}{cc}
 & \begin{array}{cccccccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \end{array} \\
\begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \end{array} &
\begin{bmatrix}
0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0
\end{bmatrix}
\end{array}
$$

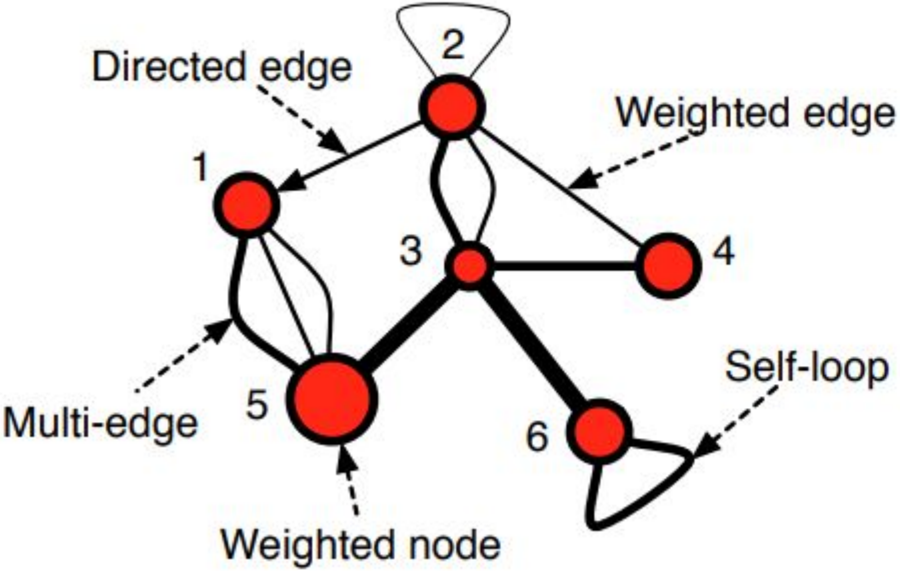
Simple and Not Simple





[From Clauset's slides]

Example



adjacency matrix

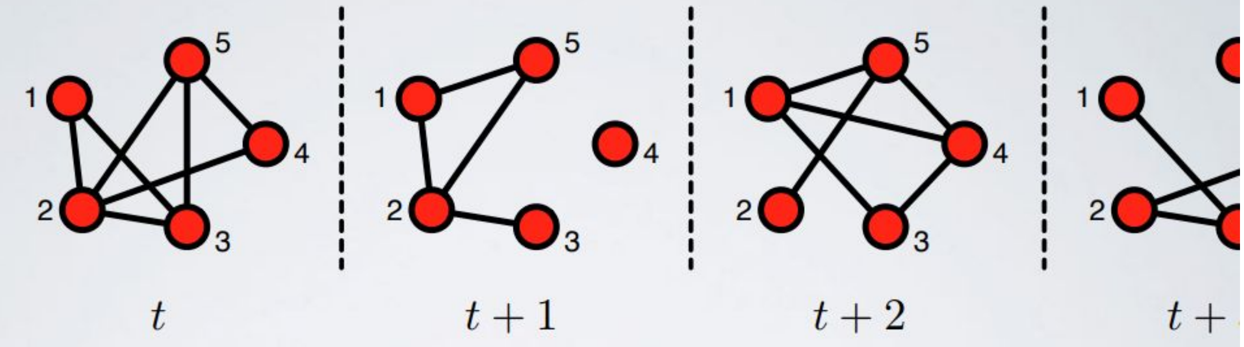
A	1	2	3	4	5	6
1	0	0	0	0	$\{1,1,2\}$	0
2	1	$\frac{1}{2}$	$\{2,1\}$	1	0	0
3	0	$\{2,1\}$	0	2	4	4
4	0	1	2	0	0	0
5	$\{1,1,2\}$	0	4	0	0	0
6	0	0	4	0	0	2

adjacency list

A	
1	$\rightarrow \{(5,1),(5,1),(5,2)\}$
2	$\rightarrow \{(1,1),(2,\frac{1}{2}),(3,2),(3,1),(4,1)\}$
3	$\rightarrow \{(2,2),(2,1),(4,2),(5,4),(6,4)\}$
4	$\rightarrow \{(2,1),(3,2)\}$
5	$\rightarrow \{(1,1),(1,1),(1,2),(3,4)\}$
6	$\rightarrow \{(3,4),(6,2)\}$

[From Clauset's slides]

Temporal Networks, snapshots or continuous



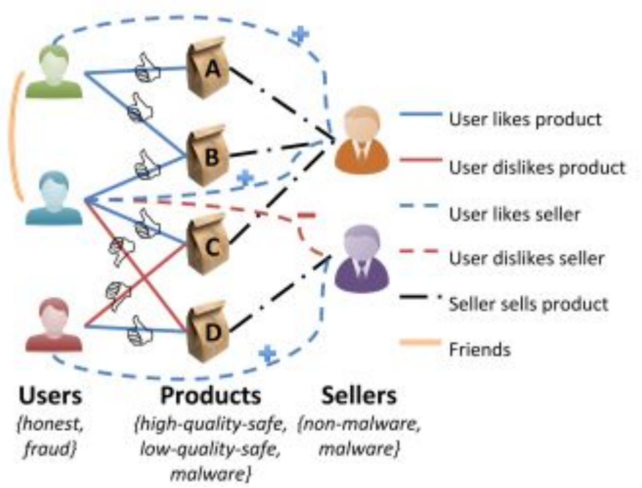
any network over time

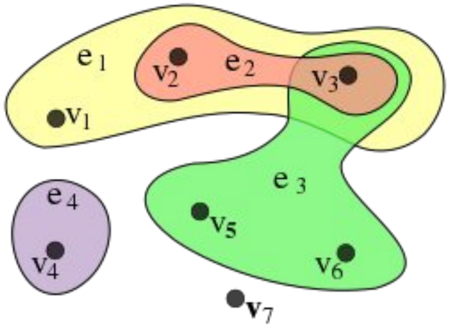
discrete time (snapshots), edges (i, j, t)

continuous time, edges $(i, j, t_s, \Delta t)$

Not Simple Graphs

- Multigraph: Multiple edges
  - E.g. followership & friendship
- Heterogeneous Graphs: Different Types
  - E.g. people, places, interest
- Relation between more than two nodes
  - Hypergraphs, E.g. family
- Relationships in different layers
  - Multiplex or multilayer network

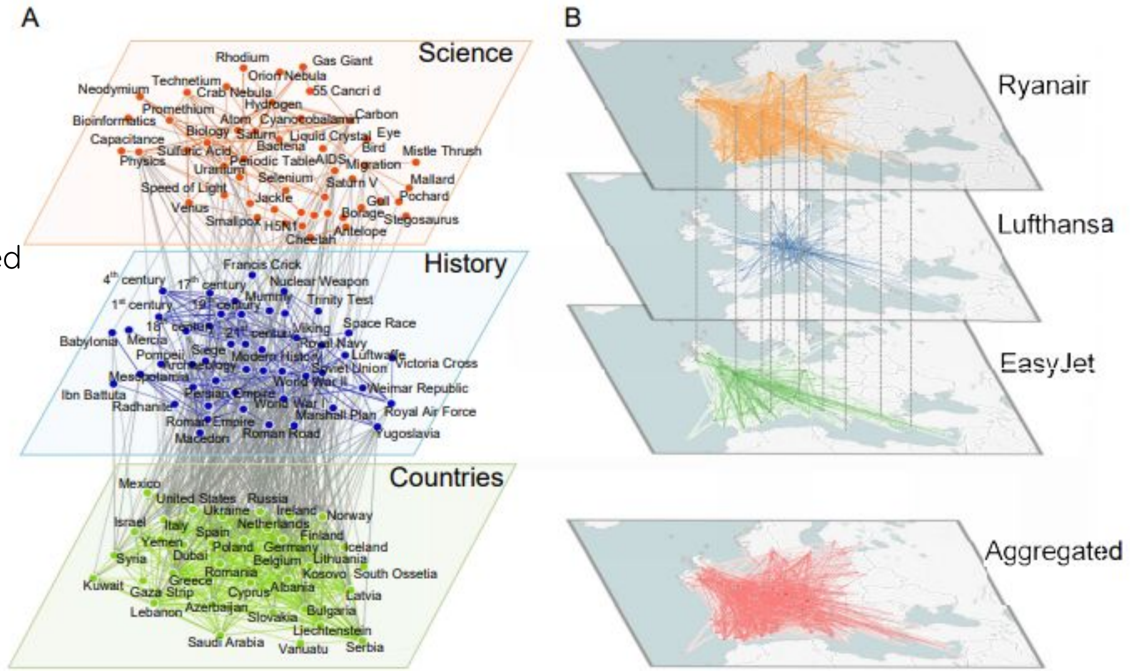




Multilayer Networks

different sets of nodes

E.g. wiki pages layered by subject



Multiplex: same set of nodes

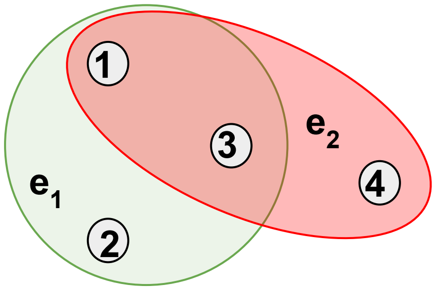
different types of connections

E.g. flights layered by airlines

https://arxiv.org/pdf/1708.07763.pdf

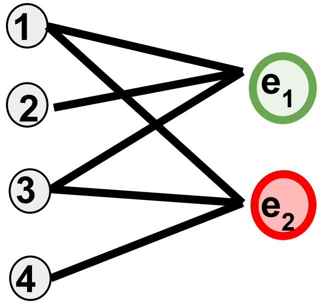
Incidence Matrix

- A_{ij}= 1 if i is connected to j & 0 otherwise
- B_{ik}=1 if i is incident to edge k & 0 otherwise
- If simple graph
  - **2 ones in each column**
  - $\mathbf{BB^T = A + D}$



B	**e_1**	**e_2**
1	1	1
2	1	0
3	1	1
4	0	1

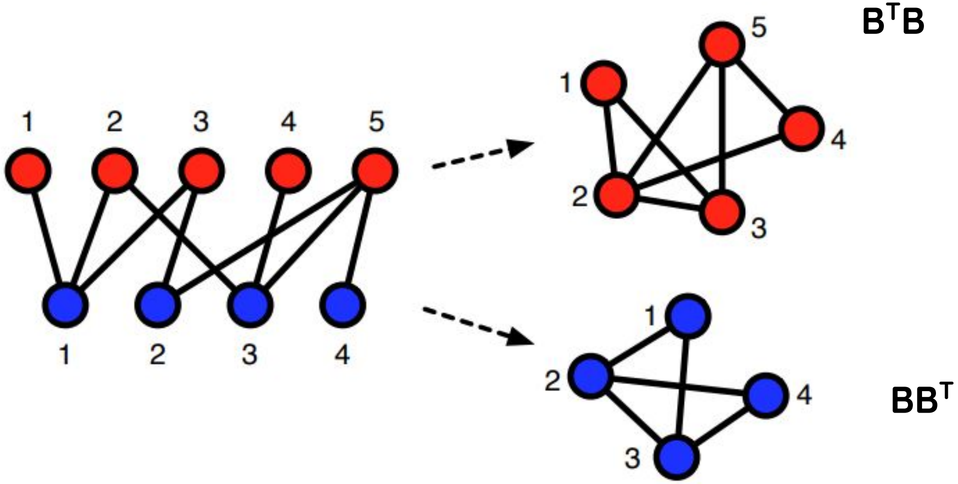
- **Bipartite Graphs**



$V = A \cup B$ where $A \cap B = \varnothing$, and $\forall (i,j) \in E((i \in A) \land (j \in B)) \lor ((i \in B) \land (j \in A))$

Bipartite Networks

$$B = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

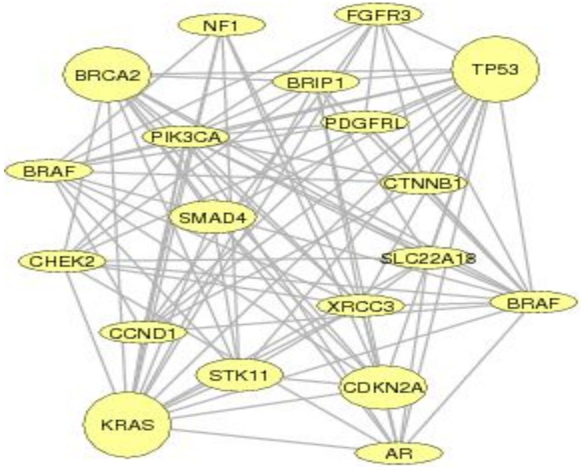


No within edges
& Two possible
One mode projections

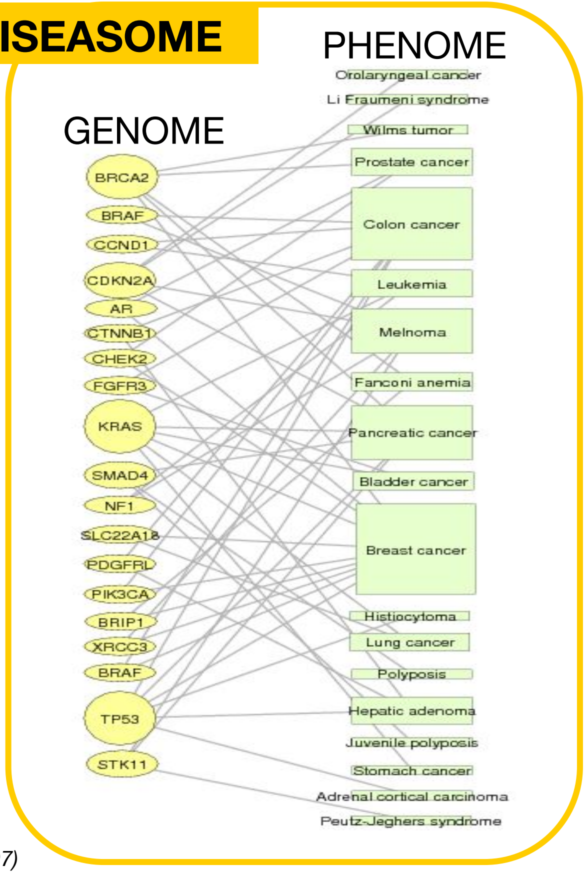
authors & papers	people & locations (checkins)
actors & movies/scenes	metabolites & reactions
musicians & albums	genes & substrings
people & online groups	words & documents
people & corporate boards	plants & pollinators

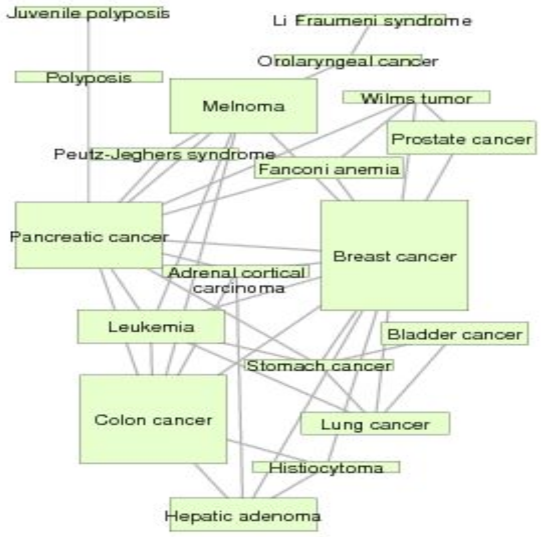
From Clauset's slides

Bipartite Networks example



Gene network



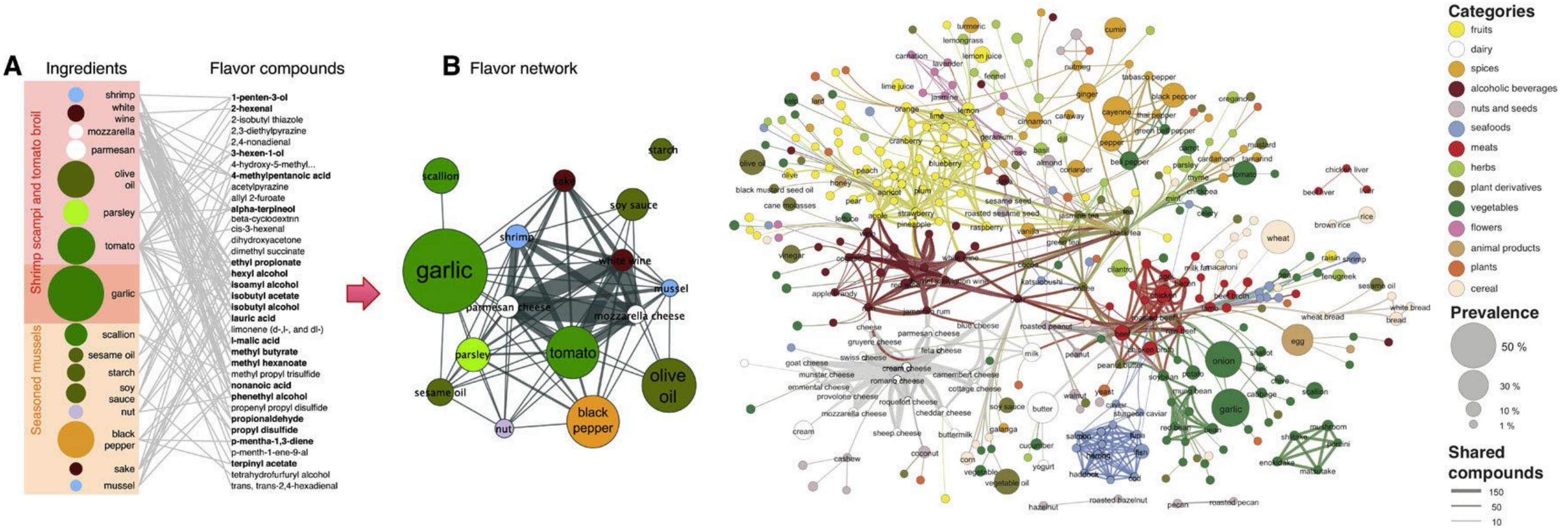


Disease network

From Barbasi's slides

Goh, Cusick, Valle, Childs, Vidal & Barabási, PNAS (2007)

Bipartite Networks example



Ingredient-Flavor Network

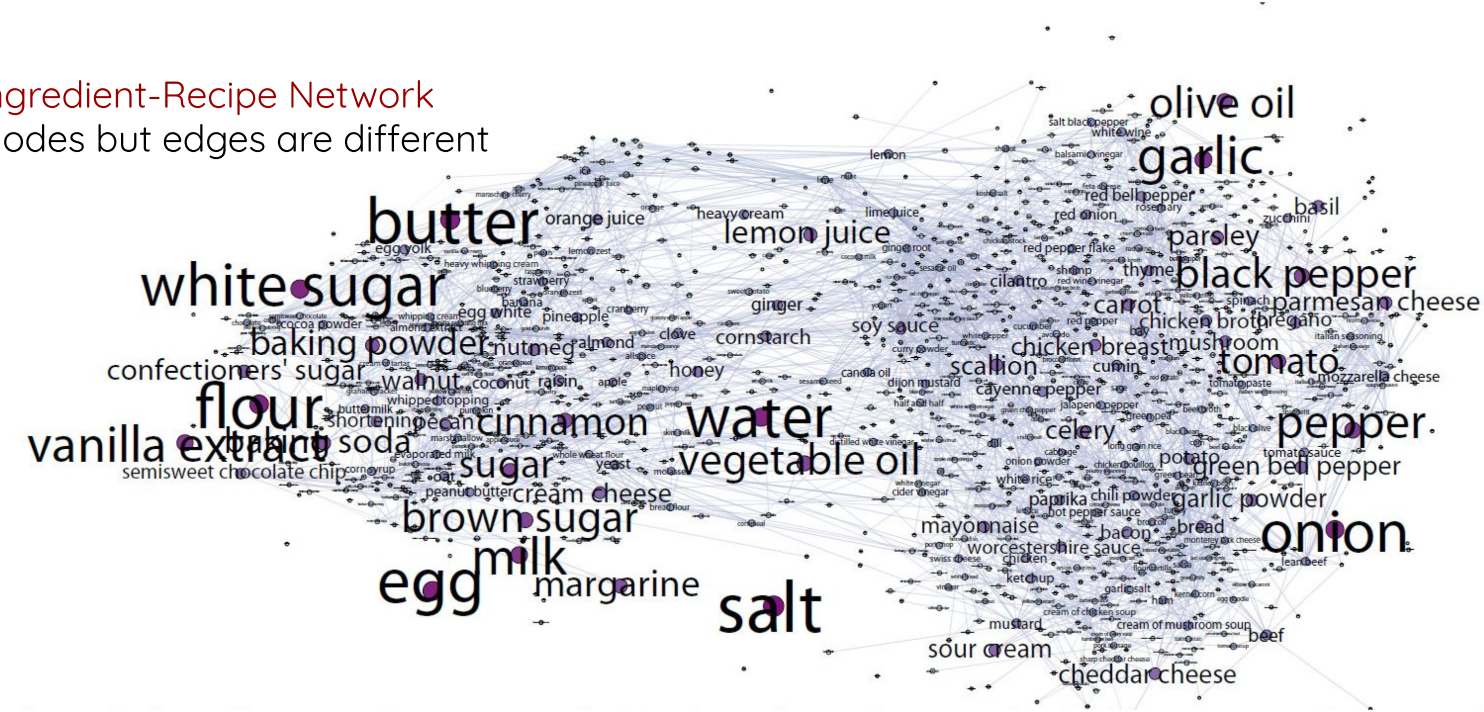
[From Barbasi's slides](#)

Y.-Y. Ahn, S. E. Ahnert, J. P. Bagrow, A.-L. Barabási Flavor network and the principles of food pairing , Scientific Reports 196, (2011).

Bipartite Networks example

From Ingredient-Recipe Network

Same nodes but edges are different



https://arxiv.org/pdf/1111.3919.pdf

https://studentwork.prattsi.org/infovis/labs/visualizing-ingredient-networks/ browse for visualizarions and project ideas