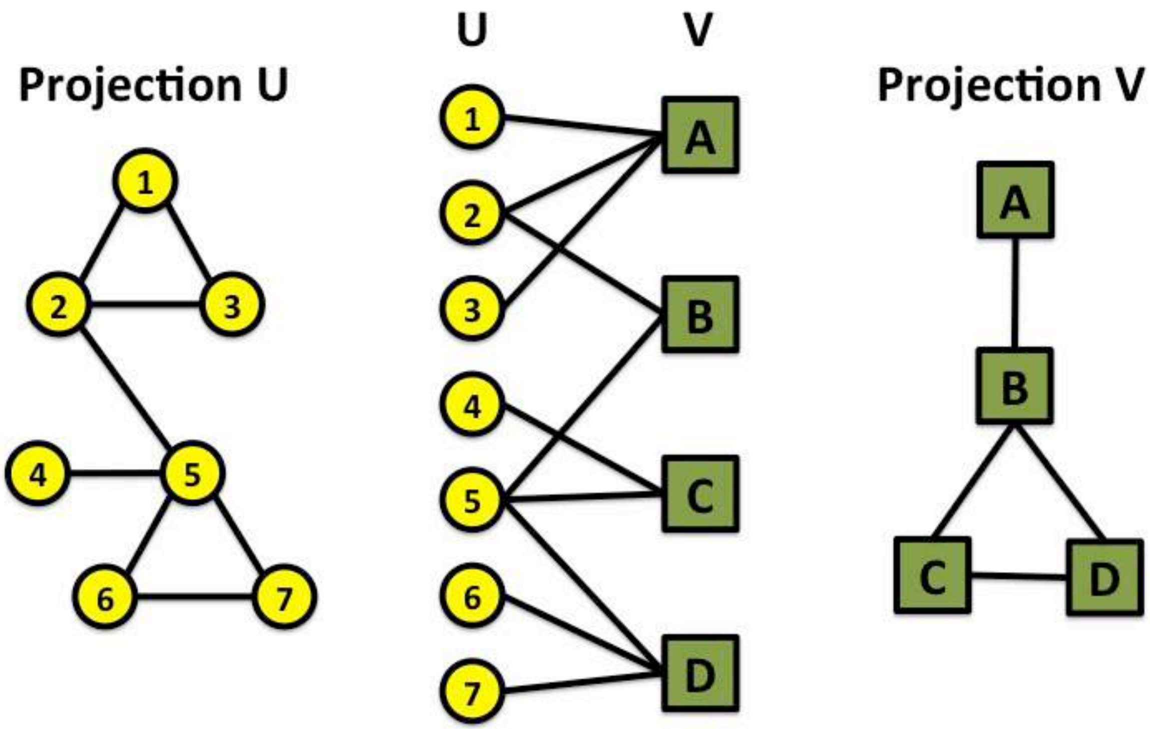
BIPARTITE NETWORKS

BIPARTITE GRAPHS

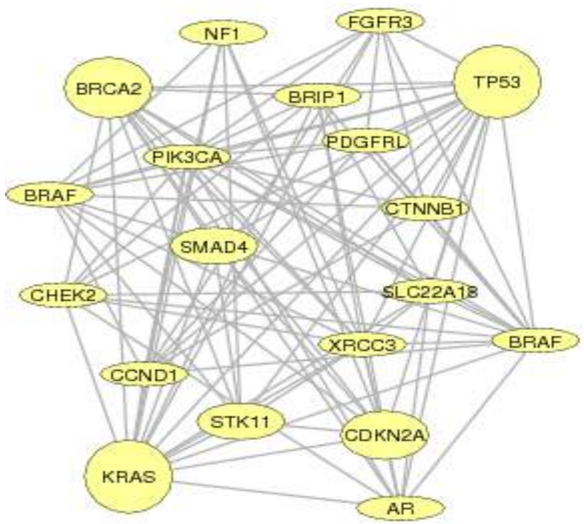
bipartite graph (or **bigraph**) is a graph whose nodes can be divided into two disjoint sets *U* and *V* such that every link connects a node in *U* to one in *V*; that is, *U* and *V* are independent sets.



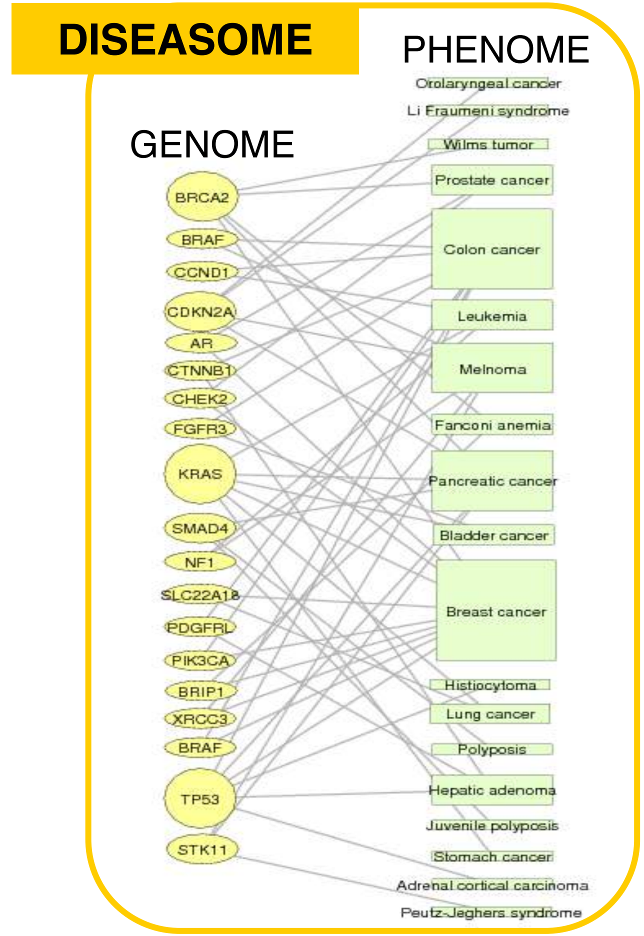
Examples:

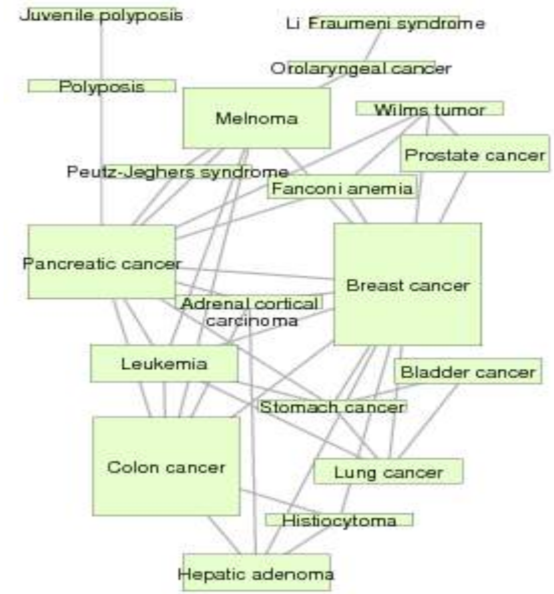
Hollywood actor network
Collaboration networks
Disease network (diseasome)

GENE NETWORK – DISEASE NETWORK



Gene network

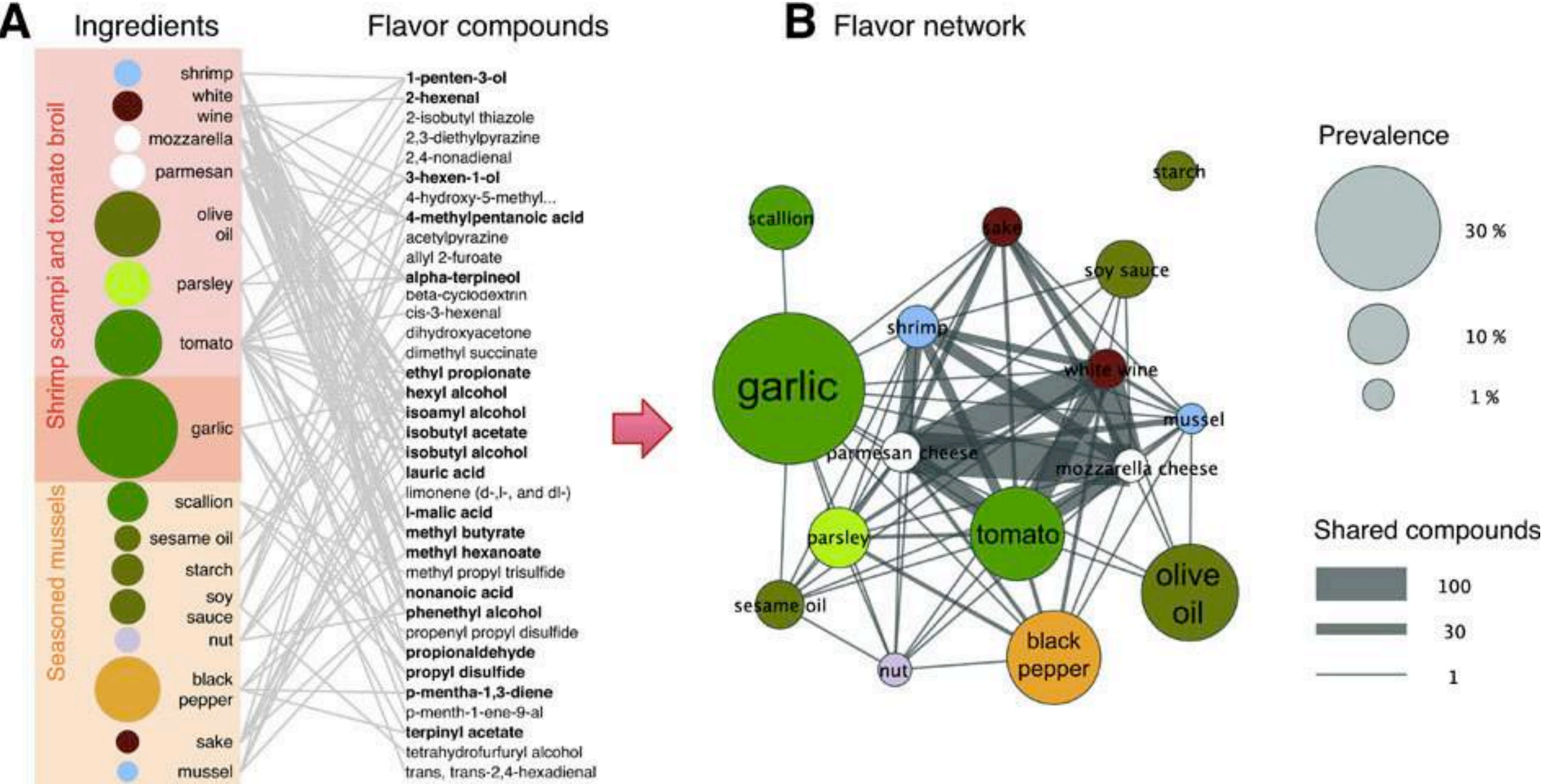




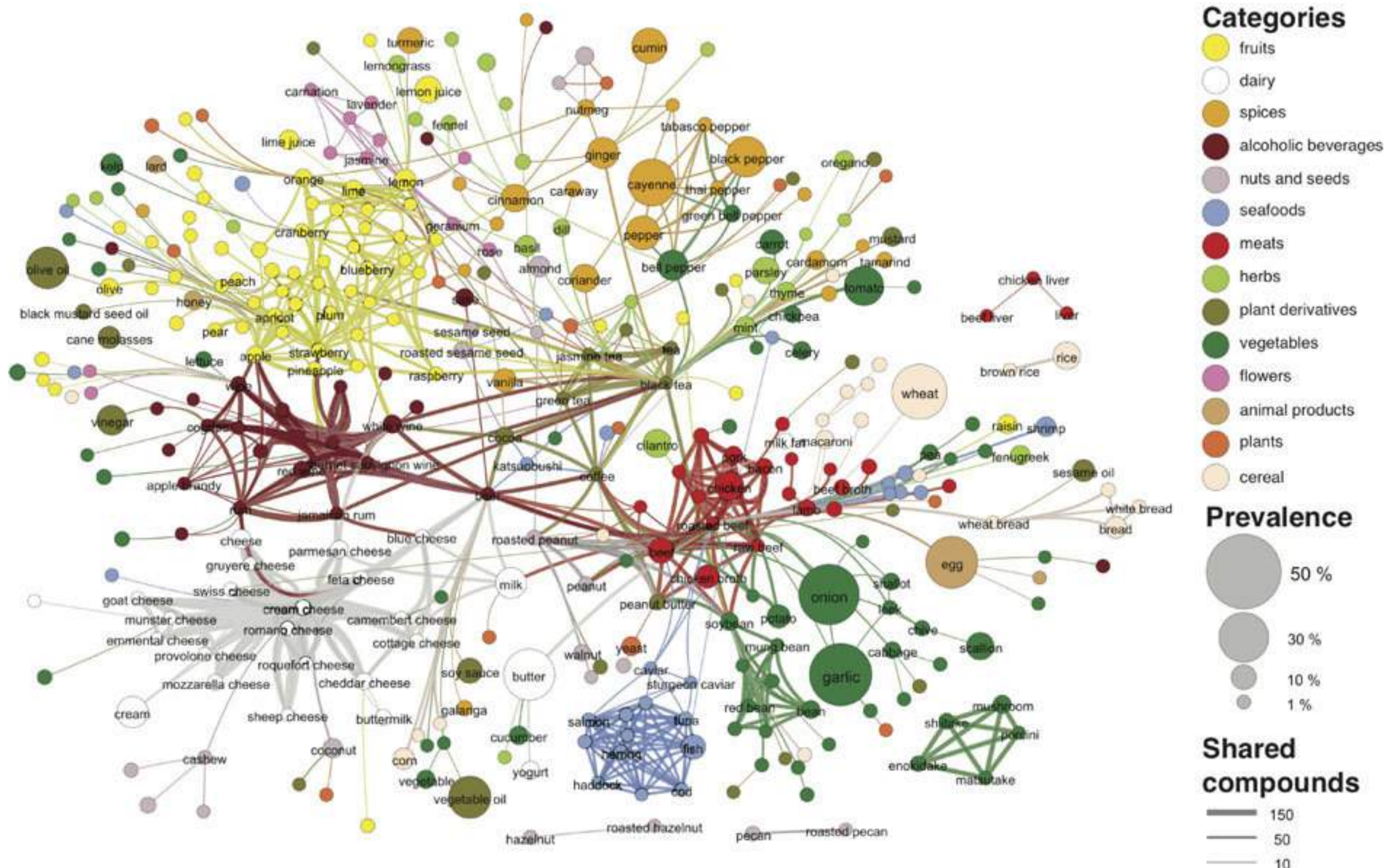
Disease network

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

Ingredient-Flavor Bipartite Network



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


Categories
fruits
dairy
spices
alcoholic beverages
nuts and seeds
seafoods
meats
herbs
plant derivatives
vegetables
flowers
animal products
plants
cereal
Prevalence
50 %
30 %
10 %
1 %
Shared compounds
150
50
10
turmeric
cumin
lemongrass
carnation
lemon juice
lavender
nutmeg
fennel
tabasco pepper
lime juice
jasmine
ginger
black pepper
oregano
kelp
lard
orange
lime
lemon
cayenne
thai pepper
caraway
cinnamon
green bell pepper
geranium
cranberry
dill
pepper
carrot
mustard
rose
basil
cardamom
tamarind
olive oil
blueberry
almond
coriander
bell pepper
parsley
tomato
chicken liver
peach
olive
thyme
honey
black mustard seed oil
apricot
plum
chickpea
beef liver
liver
sesame seed
mint
pear
cane molasses
roasted sesame seed
apple
strawberry
tea
celery
lettuce
jasmine tea
rice
pineapple
raspberry
vanilla
black tea
brown rice
wheat
green tea
vinegar
white wine
cocoa
cilantro
milk fat
macaroni
raisin
shrimp
pork
bacon
pea
fenugreek
katsuobushi
coffee
sesame oil
apple brandy
chicken
beef broth
lamb
white bread
beer
rum
jamaican rum
roasted beef
wheat bread
bread
cheese
blue cheese
roasted peanut
parmesan cheese
beef
raw beef
egg
gruyere cheese
feta cheese
chicken broth
milk
peanut
shallot
swiss cheese
goat cheese
cream cheese
onion
peanut butter
leek
munster cheese
camembert cheese
soybean
potato
romano cheese
chive
emmental cheese
cottage cheese
walnut
yeast
mung bean
cabbage
scallion
provolone cheese
roquefort cheese
caviar
soy sauce
butter
garlic
mozzarella cheese
cheddar cheese
sturgeon caviar
mushroom
red bean
bean
shiitake
cream
sheep cheese
buttermilk
galanga
salmon
tuna
porcini
cucumber
fish
cashew
coconut
corn
herring
enokidake
yogurt
vegetable
haddock
cod
matsutake
vegetable oil
roasted hazelnut
hazelnut
pecan
roasted pecan

Question 6

Q6: Paths

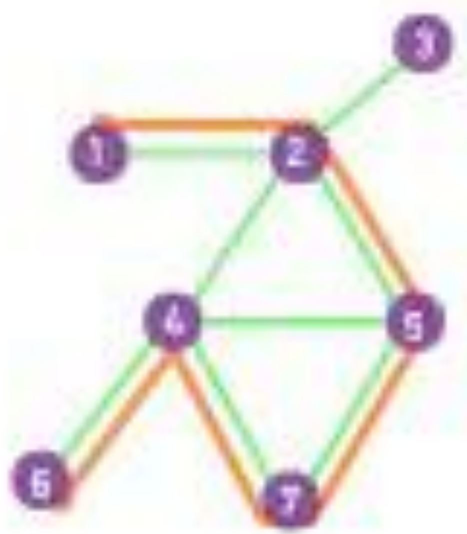
PATHOLOGY

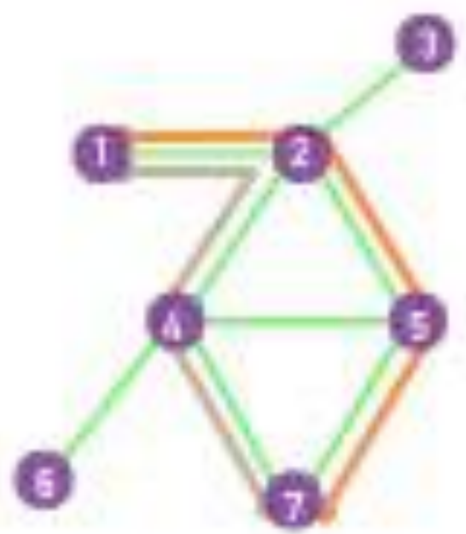
PATHS

A *path is* a sequence of nodes in which each node is adjacent to the next one

$P_{i0,in}$ of length *n* between nodes i_0 and i_n is an ordered collection of *n+1* nodes and *n* links

$$P_n = \{i_0, i_1, i_2, \ldots, i_n\} \qquad P_n = \{(i_0, i_1), (i_1, i_2), (i_2, i_3), \ldots, (i_{n-1}, i_n)\}$$

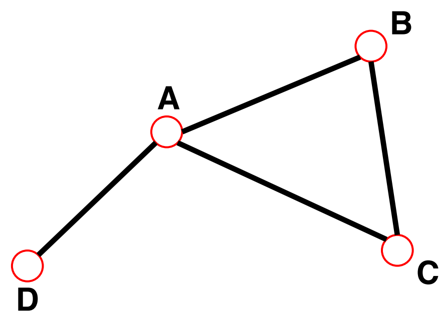




- In a directed network, the path can follow only the direction of an arrow.

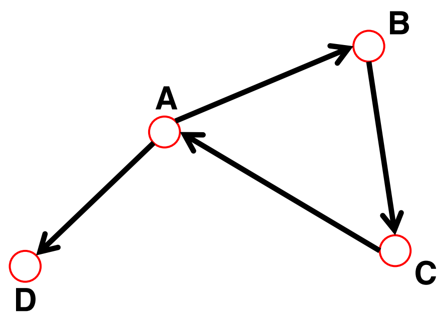
DISTANCE IN A GRAPH

Shortest Path, Geodesic Path



The *distance (shortest path, geodesic path)* between two nodes is defined as the number of edges along the shortest path connecting them.

*If the two nodes are disconnected, the distance is infinity.



In directed graphs each path needs to follow the direction of the arrows.

Thus in a digraph the distance from node A to B (on an AB path) is generally different from the distance from node B to A (on a BCA path).

N_{ij}, number of paths between any two nodes *i* and *j*:

Length n=1: If there is a link between *i* and *j*, then $A_{ij}=1$ and $A_{ij}=0$ otherwise.

Length n=2: If there is a path of length two between *i* and *j*, then $A_{ik}A_{kj}=1$, and $A_{ik}A_{kj}=0$ otherwise.

The number of paths of length 2:

$$N_{ij}^{(2)} = \sum_{k=1}^{N} A_{ik}A_{kj} = [A^2]_{ij}$$

Length n: In general, if there is a path of length *n* between *i* and *j*, then $A_{ik}\dots A_{lj}=1$ and $A_{ik}\dots A_{lj}=0$ otherwise.

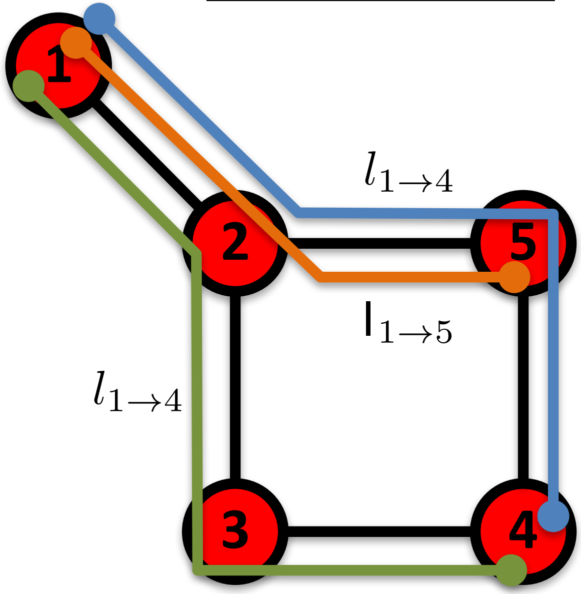
The number of paths of length *n* between *i* and *j* is*

$$N_{ij}^{(n)} = [A^n]_{ij}$$

**holds for both directed and undirected networks.*

PATHOLOGY: summary

Shortest Path



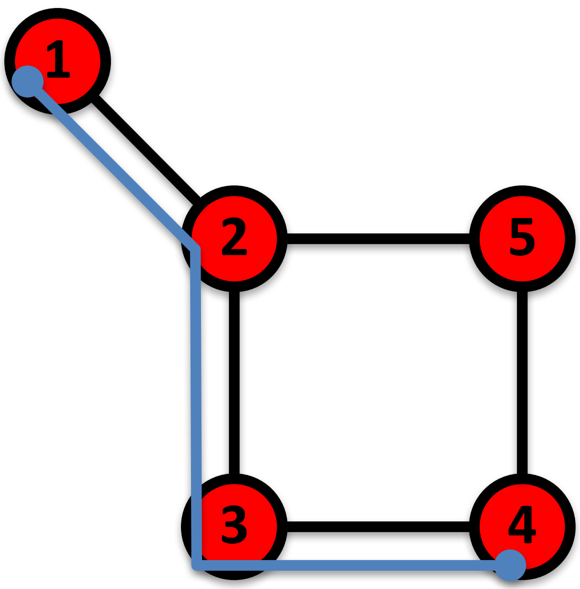
$l_{1\to4} = 3$

$l_{1\to5} = 2$

The path with the shortest length between two nodes (distance).

PATHOLOGY: summary

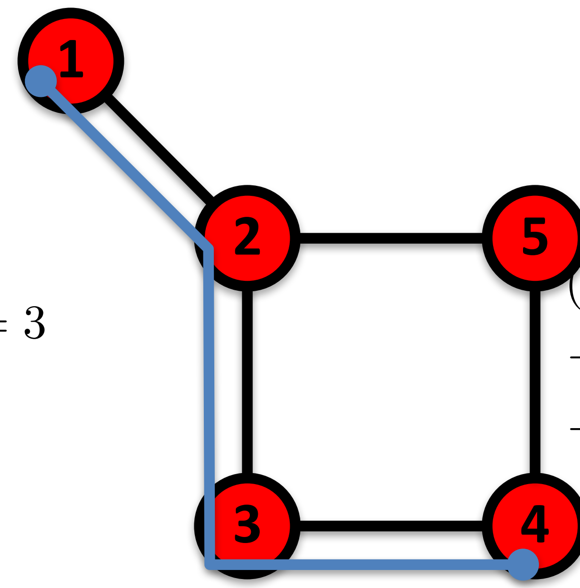
Diameter



$l_{1\to4} = 3$

The longest shortest path in a graph

Average Path Length

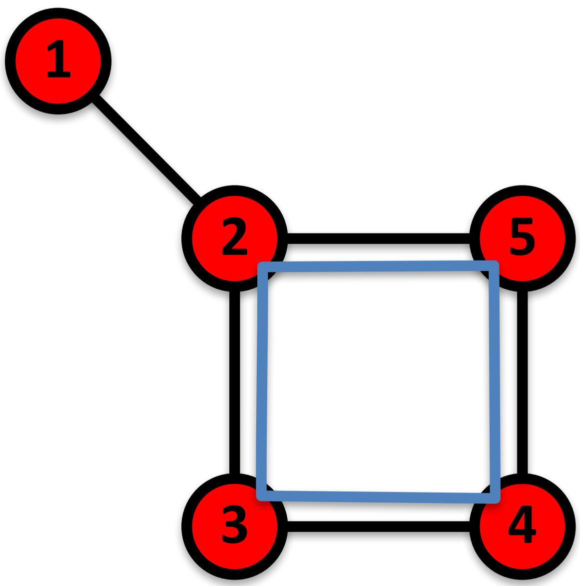


$$(l_{1\to2} + l_{1\to3} + l_{1\to4} + l_{1\to5} + l_{2\to3} + l_{2\to4} + l_{2\to5} + l_{3\to4} + l_{3\to5} + l_{4\to5})/10 = 1.6$$

The average of the shortest paths for all pairs of nodes.

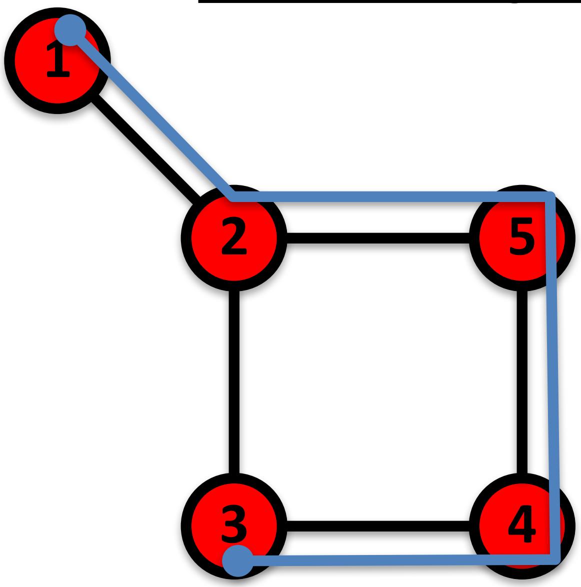
PATHOLOGY: summary

Cycle



A path with the same start and end node.

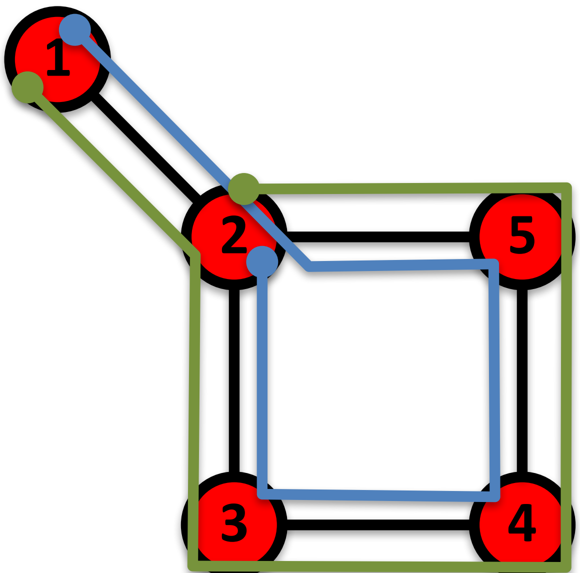
Self-avoiding Path



A path that does not intersect itself.

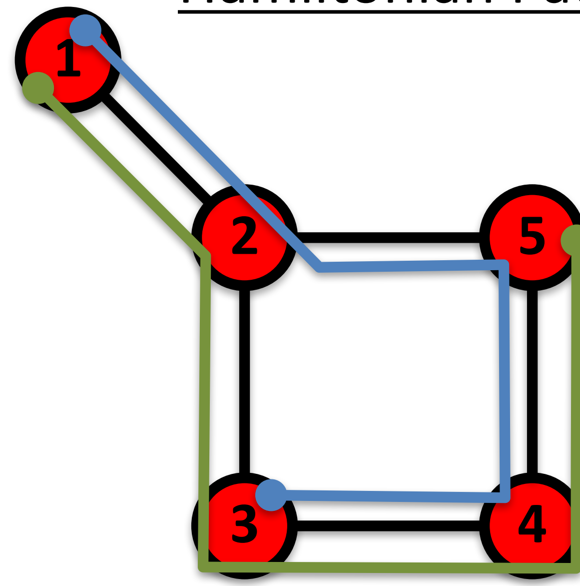
PATHOLOGY: summary

Eulerian Path



A path that traverses each link exactly once.

Hamiltonian Path

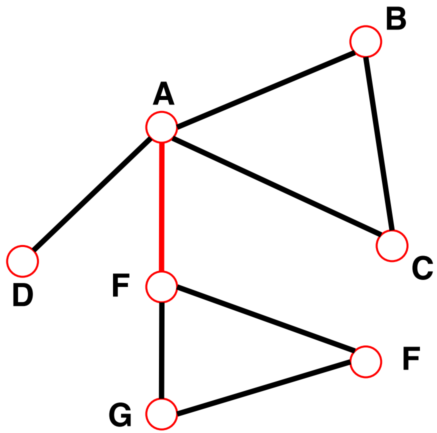


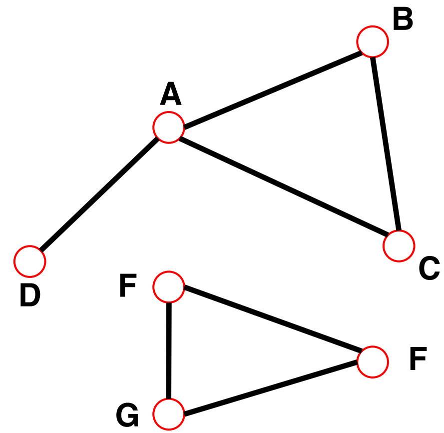
A path that visits each node exactly once.

CONNECTEDNESS

CONNECTIVITY OF UNDIRECTED GRAPHS

Connected (undirected) graph: any two vertices can be joined by a path.
A disconnected graph is made up by two or more connected components.





Largest Component:
Giant Component

The rest: **Isolates**

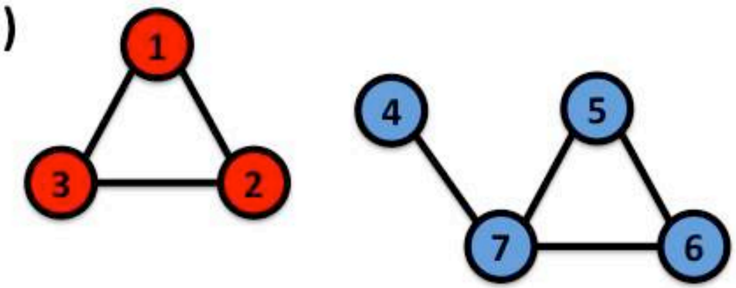
Bridge: if we erase it, the graph becomes disconnected.

CONNECTIVITY OF UNDIRECTED GRAPHS

Adjacency Matrix

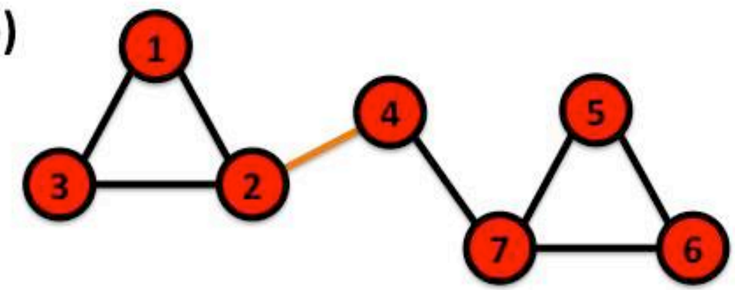
The adjacency matrix of a network with several components can be written in a block-diagonal form, so that nonzero elements are confined to squares, with all other elements being zero:

(a)



$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

(b)



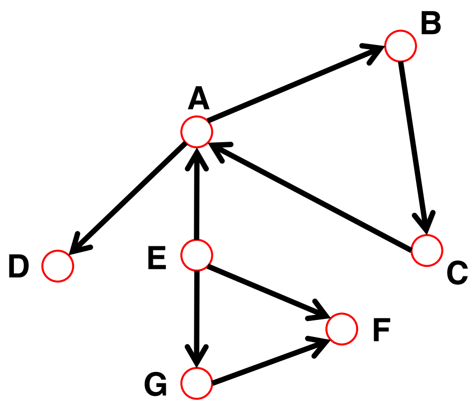
$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

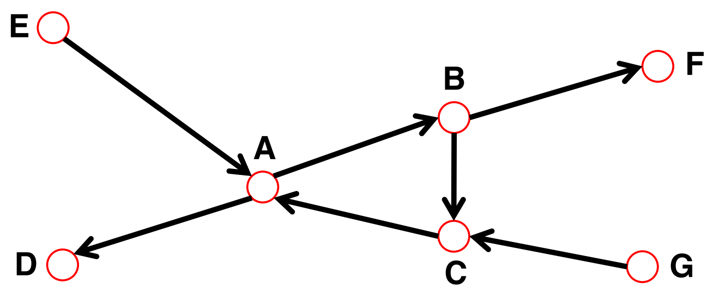
CONNECTIVITY OF DIRECTED GRAPHS

Strongly connected directed graph: has a path from each node to every other node and vice versa (e.g. AB path and BA path).

Weakly connected directed graph: it is connected if we disregard the edge directions.

Strongly connected components can be identified, but not every node is part of a nontrivial strongly connected component.





In-component: nodes that can reach the scc,

Out-component: nodes that can be reached from the scc.

FINDING THE CONNECTED COMPONENTS OF A NETWORK

1. Start from a randomly chosen node i and perform a BFS (BOX 2.5). Label all nodes reached this way with $n = 1$.

2. If the total number of labeled nodes equals N, then the network is connected. If the number of labeled nodes is smaller than N, the network consists of several components. To identify them, proceed to step 3.

3. Increase the label $n \rightarrow n + 1$. Choose an unmarked node j, label it with n. Use BFS to find all nodes reachable from j, label them all with n. Return to step 2.